目　录

第1章 导 论

进入20世纪90年代以来，人类社会正从工业经济迈向知识经济时代，一场以信息技术为核心的革命正在深刻改变着人类生活与社会的面貌，作为全球信息化浪潮重要组成部分的地理信息系统的建设与应用，日益引起科技界、企业界和政府部门的广泛关注。地理信息系统、遥感技术和全球定位技术三者有机结合，构成科学地理学日臻完善的技术体系，引起世界各国普遍的重视。地理信息系统是管理和分析空间数据的科学技术，它及时而又准确地向地学工作者、各级管理和生产部门提供有关区域综合、方案优选、战略决策等方面可靠的地理或空间信息，这就是地理信息系统的主要职能。在论述地理信息系统之前，先介绍有关基本概念。

§1 地理信息系统基本概念

1.1 数据与信息

在地理信息系统的研究和应用中，经常要涉及到数据(data)和信息(information)两个术语。从科学的观点看，两者之间有词义上的差别，即数据是信息的表达，而信息则是数据的内容。数据是未经加工的原始材料，地理信息系统的设计和建立，首先是收集数据和处理数据。有人认为，输入的都叫数据，输出的都叫信息。其实不然，数据是通过数字化或记录下来可以被鉴别的符号，不仅数字是数据，而且文字、符号和图像也是数据，数据本身并没有意义。例如数字“1”，可以离开地理信息系统而独立存在，也可以离开地理信息系统的各个组成和阶段而独立存在，即它既可以回避实体是什么，也可以回避它本身能做什么，而且在计算机化的地理信息系统中，数据的格式往往和具体的计算机系统有关，随载荷它的物理设备的形式而改变。但信息则不然，信息是对数据的解释、运用与解算，数据即使是经过处理以后的数据，只有经过解释才有意义，才成为信息。就本质而言，数据是客观对象的表示，而信息则是数据内涵的意义，只有数据对实体行为产生影响时才成为信息。例如同样的数字“1”，当用来标识某一种实体的类别时，它就提供了特征码信息；当用来表示某一种实体在某个地域内存在与否

时，它就提供了有（1表示）无（0表示）信息；当在绘图矩阵中表示绘线或不绘线时，它就提供了抬落笔信息等等。可见，信息是用数字、文字、符号、语言等介质来表示事件、事物、现象等的内容、数量或特征，以便向人们（或系统）提供关于现实世界新的事实的知识，作为生产、管理和决策的依据。

但是，要从数据中得到信息，处理和解释是非常重要的环节。所谓数据处理，是指对数据进行收集、筛选、排序、归并、转换、存储、检索、计算，以及分析、模拟和预测等等操作。这些操作的目的：①把数据转换成便于观察、分析、传输或进一步处理的形式；②把数据加工成对正确管理和决策有用的数据；③把数据编辑后存储起来，以供不断使用。数据处理是为了解释，而数据解释需要人的智慧、学识和经验。对同一数据，每个人的解释可能不同，其对决策的影响也可能不同。而不同的解释，则往往来自不同的背景、目的和应用。

信息来自数据，那么信息有些什么特点呢？首先，信息的客观性：任何信息都是与客观事实紧密相关的，这是信息的正确性和精确度的保证；其次，信息的适用性：信息对决策是十分重要的，信息系统将地理空间的巨大数据流收集、组织和管理起来，经过处理、转换和分析变为对生产、管理和决策具有重要意义的有用信息，这是由建立信息系统的明确目的性所决定的；第三，信息的传输性：信息可以在信息发送者和接受者之间传输，既包括系统把有用信息送至终端设备（包括远程终端）和以一定的形式或格式提供给有关用户，也包括信息在系统内各个子系统之间的流转和交换，目前在不少系统中，已较为广泛地采用了网络传输技术；第四，信息的共享性：信息与实物不同，信息可以传输给多个用户，为多个用户共享，而其本身并无损失，这为信息的并发应用提供可能性。信息的这些特点，使信息成为当代社会发展的一项重要资源，它已渗透到各个学科领域，在地学领域内，它使地理学与信息科学相结合，诞生了以信息论、控制论、计算机科学、系统工程和人工智能等科学为基础的新的边缘技术学科——地理信息系统。

1.2　地理信息与地理信息系统

1. 地理信息

地理信息是指表征地理圈或地理环境固有要素或物质的数量、质量、分布特征、联系和规律等的数字、文字、图像和图形等的总称。从地理实体到地理数据、从地理数据到地理信息的发展，反映了人类从认识物质、能量到认识信息的一个巨大飞跃。地理圈或地理环境是客观世界最大的信息源，随着现代科学技术的发展，特别是借助于近代数学、空间科学和计算机科学，科学工作者已经有可能迅速地采集到地理空间的几何信息、物理信息和人为信息，并定期和适时地识别、转换、存储、传输、显示和控制应用这些信息，这也已经成为现代地理学的重要任务之一。

地理信息属于空间信息,其位置的识别是与数据联系在一起的,这是地理信息区别于其他类型信息的一个最显著的标志。地理信息的这种定位特征是通过公共的地理基础来体现的,即按照特定地区的经纬网或公里网建立的地理坐标来实现空间位置的识别,并可以按照指定的区域进行信息的并或分。其次,地理信息具有多维结构的特征,即在二维空间的基础上,实现多专题的第三维的信息结构,而各个专题型或实体型之间的联系是通过属性码进行的,这既对岩石圈-气圈-水圈-生物圈及其内部的相互作用进行综合性的研究提供了可能性,也为地理圈多层次的分析和信息的传递与筛选提供了方便。第三,地理信息的时序特征十分明显,因此可以按照时间的尺度进行地理信息的划分,分为超短期的(如台风、森林火灾)、短期的(如江河洪水、作物长势)、中期的(如土地利用、作物估产)、长期的(如水土流失、城市化)和超长期的(如火山爆发、地壳变形)等。地理信息的这种动态变化的特征,一方面要求信息获取及时、定期更新,另一方面要重视自然历史过程的积累和对未来的预测、预报,以免用过时的信息造成决策的失误,或者缺乏可靠的动态数据,不能对事变中的地理事件或现象作出合乎机理的预测预报和科学论证。因此,要研究地理信息,首先必须把握地理信息的这种区域性的、多层次的和动态变化的特征,然后才能选择正确的手段,实现人口、资源和环境的综合分析、管理、规划和决策。

2. 地理信息系统

地理信息系统简称为 GIS,关于它确切的全称,多数人认为是 Geographical Information System,也有人认为是 Geo-information System。国际上现发行的两种主要的专业杂志,就是各自采用不同的全称,前者是英国出版的季刊的全称,后者是德国出版的季刊的全称。在加拿大和澳大利亚,则称为 Land Information System。在我国,通常称为 Resources and Enviromental Information Systems。全称虽各有出入,但简称都叫做 GIS。

那么,什么是 GIS 呢?对于不同的部门和不同的应用目的,其定义也不尽相同。例如,美国学者 Parker 认为"GIS 是一种存储、分析和显示空间与非空间数据的信息技术"。Goodchild 把 GIS 定义为"采集、存储、管理、分析和显示有关地理现象信息的综合系统"。加拿大的 Reger Tomlinson 认为"GIS 是全方位分析和操作地理数据的数字系统"。Burrough 认为"GIS 是属于从现实世界中采集、存储、提取、转换和显示空间数据的一组有力的工具"。俄罗斯学者也把 GIS 定义为"一种解决各种复杂的地理相关问题,以及具有内部联系的工具集合"。这些定义,有的侧重于 GIS 的技术内涵,有的则是强调 GIS 的应用功能。为了能更具体地认识和真正了解 GIS 的概念,作者推荐美国联邦数字地图协调委员会(FICCDC)关于 GIS 的定义及概念框架(图 1-1)。该定义认为"GIS 是由计算机硬件、软件和不同的方法组成的系统,该系统设计支持空间数据的采集、管理、处理、分

析、建模和显示，以便解决复杂的规划和管理问题"。根据这个定义及它的概念框架，可得出 GIS 的如下基本概念：

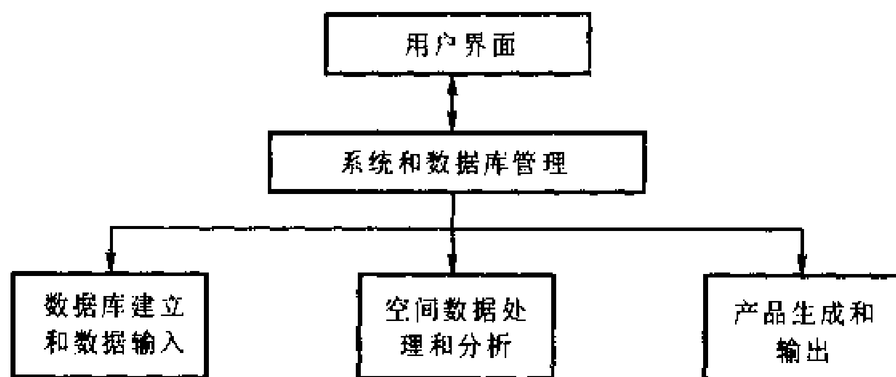


图 1-1 GIS 概念框架和构成

(1) GIS 的物理外壳是计算机化的技术系统，该系统又由若干个相互关联的子系统构成，如数据采集子系统、数据管理子系统、数据处理和分析子系统、可视化表达与输出子系统等，这些子系统的构成直接影响着 GIS 的硬件平台、系统功能和效率、数据处理的方式和产品输出的类型。

(2) GIS 的操作对象是空间数据，即以点、线、面方式编码并以(x,y)坐标串储存管理的离散型空间数据，或者以一系列栅格单元表达的连续型空间数据。空间数据的最根本特点是每一个地理目标都按统一的地理坐标进行编码，实现对其定位、定性、定量和拓扑关系的描述。GIS 以空间数据作为处理和操作的主要对象，这是它区别于其他类型信息系统的根本标志，也是其技术难点之所在。

(3) GIS 的技术优势在于它的混合数据结构和有效的数据集成、独特的地理空间分析能力、快速的空间定位搜索和复杂的查询功能、强大的图形创造和可视化表达手段，以及地理过程的演化模拟和空间决策支持功能等。其中，通过地理空间分析可以产生常规方法难以获得的重要信息，实现在系统支持下的地理过程动态模拟和决策支持，这就是 GIS 的研究核心，也是 GIS 的重要贡献。

(4) GIS 与地理学和测绘学有着密切的关系。地理学是一门研究人地相互关系的科学，研究各自然界面的生物、物理、化学过程，以及探求人类活动与资源环境间相互协调的规律，这为 GIS 提供了有关空间分析的基本观点与方法，成为 GIS 的基础理论依托。测绘学不但为 GIS 提供各种不同比例尺和精度的定位数据，而且其理论和算法可直接用于空间数据的变换和处理。而 GIS 引入地学界，正如美国地质学家 K.I.兰菲尔所说的"GIS 引入地学界，如同 Fortran 语言引入计算机科学界一样重要"，GIS 是以一种全新的思想和手段来解决复杂的规划、管理和地理相关问题，例如城市规划、商业选址、环境评估、资源管理、灾害监测、全球变化，甚至在现代企业中作为制定科学经营战略的一种重要手段，因为企业对

外界的认知能力和信息处理能力提高了,就能创造空间上的竞争优势。解决这些复杂的空间规划和管理问题,是GIS应用的主要目标。

地理信息系统根据其研究范围,可分为全球性信息系统和区域性信息系统;根据其研究内容,可分为专题信息系统和综合信息系统;根据其使用的数据模型,可分为矢量、栅格和混合型信息系统(图1-2)。

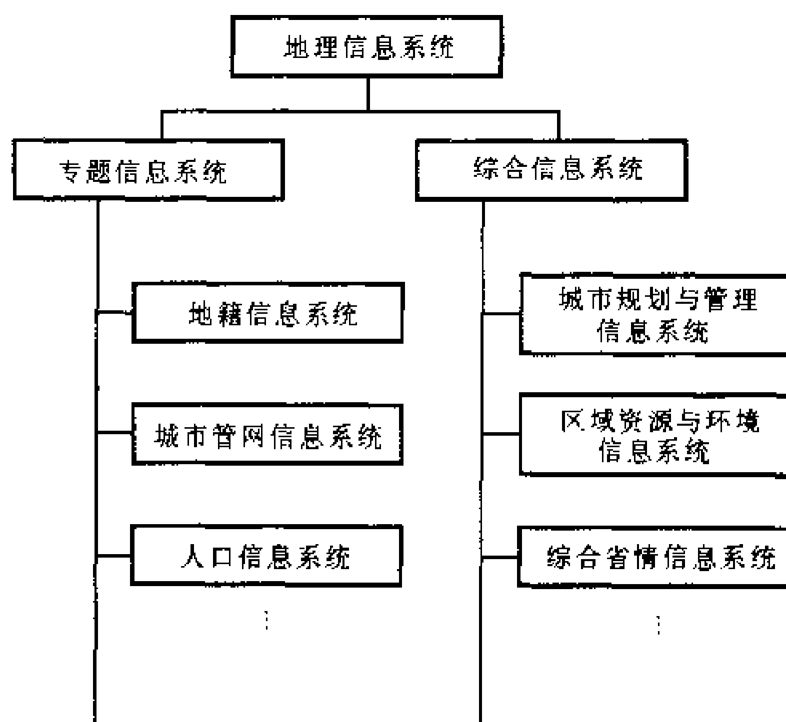


图1-2　地理信息系统的分类

§2　GIS的基本构成

一个实用的GIS,要支持对空间数据的采集、管理、处理、分析、建模和显示

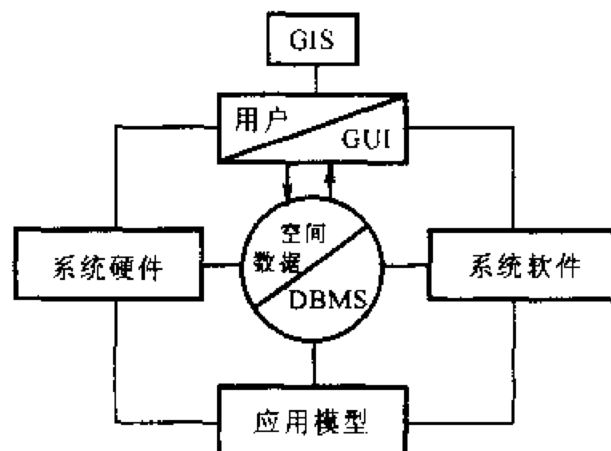


图1-3　地理信息系统构成示意图

等功能，其基本构成一般包括以下5个主要部分：系统硬件、系统软件、空间数据、应用人员和应用模型。它们之间的关系如图1-3所示。

2.1　系统硬件

GIS的硬件平台的基本类型如图1-4所示，用以存储、处理、传输和显示地理信息或空间数据。计算机与一些外部设备及网络设备的联接构成GIS的硬件环境。计算机是GIS的主机，它是硬件系统的核心，包括从主机服务器到桌面工作站，用作数据的处理、管理与计算。GIS外部设备包括输入设备的数字化仪、扫描仪和全站型测量仪器等，输出设备的绘图仪、打印机和高分辨率显示装置等，以及数据存储与传送设备的磁带机、光盘机、活动硬盘和硬盘阵列等。GIS的网络设备包括布线系统、网桥、路由器和交换机等，具体的网络设备根据网络计算的体系结构来确定。

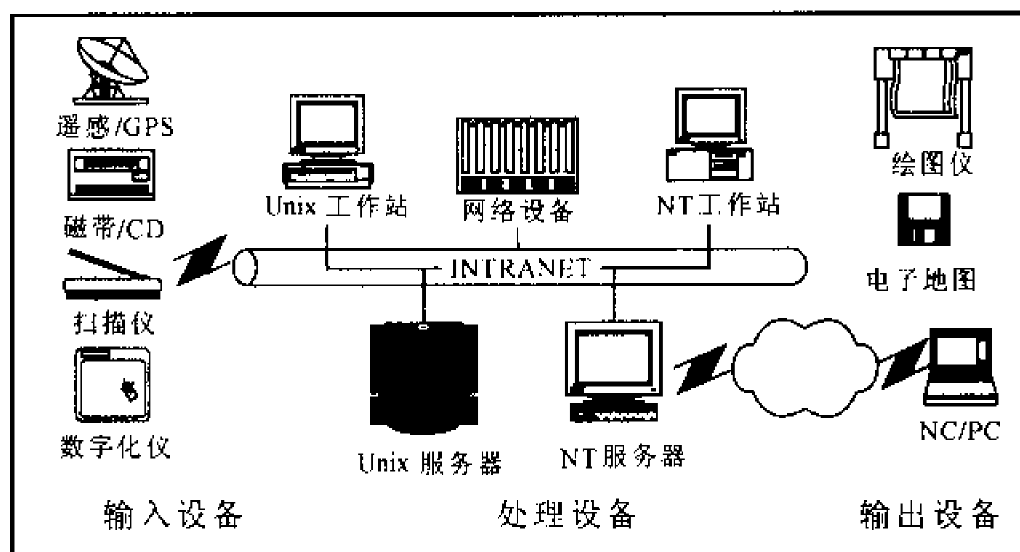


图1-4　GIS的硬件配置

1. GIS主机

目前运行GIS的主机，包括大型、中型、小型机，工作站/服务器和微型计算机，其中各种类型的工作站/服务器成为GIS的主流，特别是由Intel硬件和Windows NT构成的PC工作站正成为工作站市场的新宠，传统Unix阵营的用户正在逐渐向它转移。NT工作站对GIS用户的吸引力，包括相对低成本、可管理性、标准图形化平台和具有PC结构与效率等，因此广泛应用于GIS和某些科学应用领域。例如，ARC/INFO、Intergraph、MapInfo和GenaMap等主流GIS产品，都相继开发出其NT版本，但目前功能与Unix版相比仍有待提升。

服务器作为在网络环境下提供资源共享的主流计算产品，具有可靠性、高性能、高吞吐能力、大内存容量等特点，具备强大的网络功能和友好的人机界面，是

以网络为中心的 GIS 和现代计算环境的关键,其中以低价格和高性能为特点的 PC 服务器,正在迅速缩小与 Unix 服务器之间的差距,日益引起 GIS 设计者和用户的广泛关注。

目前 GIS 工作站和服务器主要有 Unix 和 NT 两大类型,其产品包括 SUN、HP、IBM、SGI 和 COMPAQ 等,不同种类机型的界线逐渐模糊。由于客户/服务器环境的流行,多媒体技术的发展,以及计算机与通讯技术的融合,促使 GIS 向不依赖于平台的方向发展,GIS 软件标准逐渐统一。

2. GIS 外部设备

GIS 外部设备主要包括各种输入和输出设备。主要的输入设备有图形数字化仪、图形扫描仪、解析和数字摄影测量设备等。图形数字化仪(图 1-5)尽管成本高、工序繁琐、对操作人员素质要求较高,但至今仍为空间数据采集的主要方式。市场上出售的数字化仪,例如 Calcomp Drawing-Board 系列数字化仪,其有效面积范围从 305×305 mm 到 1 118×1 524 mm,有多种配置可供选择,如不透光的或带背光的板面,有线或无线的,笔式或鼠标式定标器,和所有的应用软件都能兼容,用户可用命令设置数字化板的菜单和定标器的按键,操作模式有:Prompt Point, Run, Line, Track, Increment, Mouse, Delta, Gridupdate 等。图形数字化仪由电磁感应板、游标和相应的电子电路组成。当使用者在电磁感应板上移动游标的十字丝交点对准指定图形的点位时,按动相应的按钮,数字化仪便将对应的命令符号和该点的坐标(x, y),通过接口(多用串行接口)电路传送给计算机,定位点的精度可达 0.13 mm~0.025 mm。

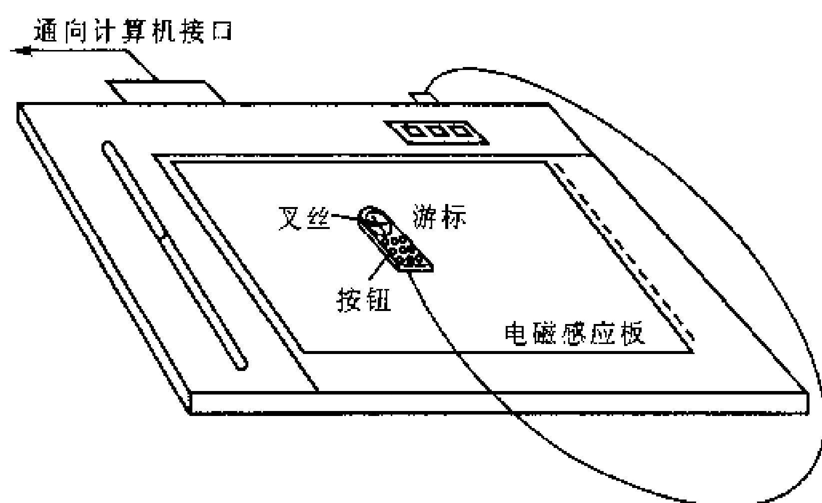


图 1-5 图形数字化仪

手扶跟踪数字化仪的速度慢,工作效率较低,而栅格数据的获取则相对容易得多,目前新一代大幅面图形扫描仪提供高分辨率、真彩色、近乎完美的图像效果,其中 ANATech 公司为用户提供了一整套应用于各领域的高精度大幅面扫描

仪,以及相配套应用软件,是图形、图像数据录入和采集最有效的工具之一。ANATech 系列扫描仪包括:

- Evolution SLI3840 高速大幅面扫描仪;
- Evolution 4080ET 超大幅面高分辨率扫描仪;
- Evolution 4080C 超大幅面高分辨率彩色扫描仪;
- Evolution 4240 超大幅面扫描仪;
- Evolution 4240C 超大幅面彩色扫描仪。

Evolution 3840 大幅面扫描仪扫描一幅 A_0 图纸时间仅需 15 s,精度为 0.05%,失真率小于 0.1%,用户可在 800 dpi 范围内任选扫描分辨率,可以按黑白二值或 256 级灰度方式扫描,扫描同时可以边显示边扫描,并具有实时消蓝去污功能 ,根据用户需求,可以自动补线、校正、镜像、反转等功能。地图扫描数字化得到图像信息,再经过目标识别和由栅格到矢量的转换过程,如图 1-6 所示。

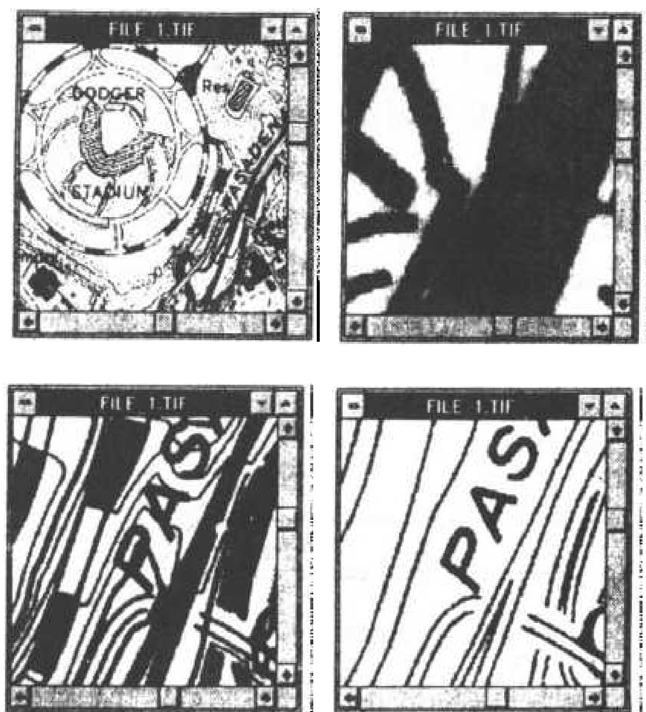


图 1-6 地图的扫描数字化

主要的输出设备有各种绘图仪、图形显示终端和打印机等。绘图仪如 HP DesignJet 750C 彩色喷墨绘图仪,是一种快速、可靠,便于联网且可在多种介质上进行高质量输出的绘图仪,是目前广泛使用的主流 GIS 产品输出设备。它采用根据对象空间分布形式和输出产品的特征,选择恰当的图形表示方法,结合色彩、线条、符号、文字等表示手段,具有 600 dpi 分辨率的高精度黑白输出,彩色输出在 300 dpi 时,颜色可多达 1 600 多种,可获得极高清晰度的绘图质量。

图形显示终端用于图形的交互式输入、编辑、分析、处理和输出。目前有多

种系列、型号的显示终端,如 Tektronix 公司生产的 4128、4335 型,与 IBM - PC 兼容的 SGS - 430 三维转换器等。

GIS 还有多种表格、文字的数据需要输出,可利用多种打印机完成。打印机的类型有针式打印机、激光打印机、液晶打印机等。HP Design Jet 彩色打印机还能打印出丰富绚丽的彩色图形和细腻的文字。

3. GIS 网络设备

进入 20 世纪 90 年代以来,计算机技术的飞速发展不断改变着 GIS 的体系结构,从主机—终端到 Client/Server 再到 Internet/Intrant。目前,基于客户/服务器体系结构,并在局域网、广域网或因特网支持下的分布式系统结构(图 1 - 7)已经成为 GIS 硬件系统的发展趋势。因此网络设备和计算机通信线路的设计成为 GIS 硬件环境的重要组成部分。

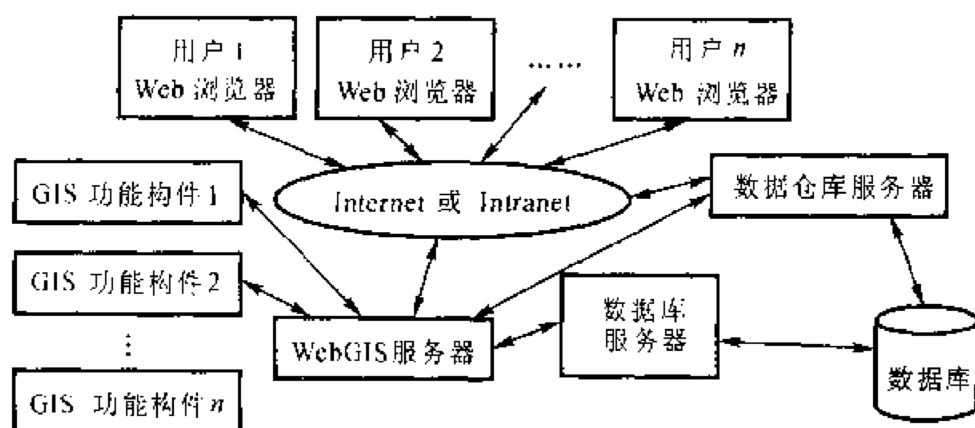


图 1 - 7 分布式 GIS 的体系结构

网络设备包括布线系统、网桥、路由器和交换机等。在进行 GIS 网络设计时,必须首先确定网络应用的需求,然后具体考虑网络类型、互联设备、网络操作系统和服务器的选择,以及网络拓扑结构、网络布线和网络安全性保障等。只有对新技术的深刻理解、对新产品的广泛关注以及对应用需求的准确把握,才能设计出一个合理的 GIS 网络。

2.2 系统软件

GIS 软件是系统的核心,用于执行 GIS 功能的各种操作,包括数据输入、处理、数据库管理、空间分析和图形用户界面(GUI)等。按照其功能分为:GIS 专业软件,数据库软件和系统管理软件等(图 1 - 8)。

1. GIS 专业软件

GIS 专业软件一般指具有丰富功能的通用 GIS 软件,它包含了处理地理信息的各种高级功能,可作为其他应用系统建设的平台。其代表产品有

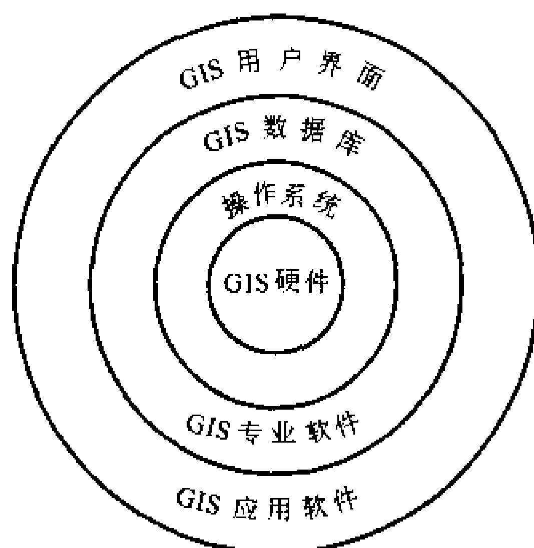


图1-8 GIS的软件层次

ARC/INFO、Mge、MapInfo、MapGIS、GeoStar等。它们一般都包含有以下的主要核心模块：

(1) 数据输入和编辑。支持手扶跟踪数字化、图形扫描及矢量化，以及对图形和属性数据提供修改和更新等编辑操作。

(2) 空间数据管理。能对大型的、分布式的、多用户数据库进行有效的存储、检索和管理。

(3) 数据处理和分析。能转换各种标准的矢量格式和栅格格式数据，完成地图投影转换，支持各类空间分析功能等。

(4) 数据输出。提供地图制作、报表生成、符号生成、汉字生成和图像显示等。

(5) 用户界面。提供生产图形用户界面工具，使用户不用编程就能制作友好和美观的图形用户界面。

(6) 系统二次开发能力。利用提供的应用开发语言，可编写各种复杂的 GIS 应用系统。

2. 数据库软件

数据库软件除了在 GIS 专业软件中用于支持复杂空间数据的管理软件以外，还包括服务于以非空间属性数据为主的数据库系统，这类软件有：Oracle，Sybase，Informix，DB2，SQL Server，Ingress 等。它们也是 GIS 软件的重要组成部分，而且由于这类数据库软件具有快速检索、满足多用户并发和数据安全保障等功能，目前已实现在现成的关系型商业数据库中存储 GIS 的空间数据，例如 SDE(Spatial Datalaee Engine)就是最好的解决方案。

3. 系统管理软件

系统管理软件主要指计算机操作系统，当今使用的操作系统有：MS－DOS，Unix，Windows，Windows 95/98，Windows NT，VMS 等。它们关系到 GIS 软件和开发语言使用的有效性，因此也是 GIS 软硬件环境的重要组成部分。

2.3 空间数据

地理信息系统的操作对象是空间数据，它具体描述地理实体的空间特征、属性特征和时间特征。空间特征是指地理实体的空间位置及其相互关系；属性特征表示地理实体的名称、类型和数量等；时间特征指实体随时间而发生的相关变化。根据地理实体的空间图形表示形式，可将空间数据抽象为点、线、面三类元素，它们的数据表达可以采用矢量和栅格两种组织形式，分别称为矢量数据结构和栅格数据结构。

在地理信息系统中，空间数据是以结构化的形式存储在计算机中的，称为数据库。数据库由数据库实体和数据库管理系统组成。数据库实体存储有许多数据文件和文件中的大量数据，而数据库管理系统主要用于对数据的统一管理，包括查询、检索、增删、修改和维护等。

由于 GIS 数据库存储的数据包含空间数据和属性数据，它们之间具有密切的联系，因此如何实现两者之间的连接、查询和管理，这是 GIS 数据库管理系统必须解决的重要问题。目前采用的解决方法有以下三种方式：

1．混合式

混合式数据库管理系统是利用两个子系统分别存储空间数据和属性数据，其中空间数据存储在线状或面状实体的弧段文件中，属性数据存储在关系数据库管理系统中，两个子系统之间通过标识码（ID）进行连接（图 1－9）。属于这种管理方式的 GIS 软件有 ARC/INFO、Mge、Sicad 等。

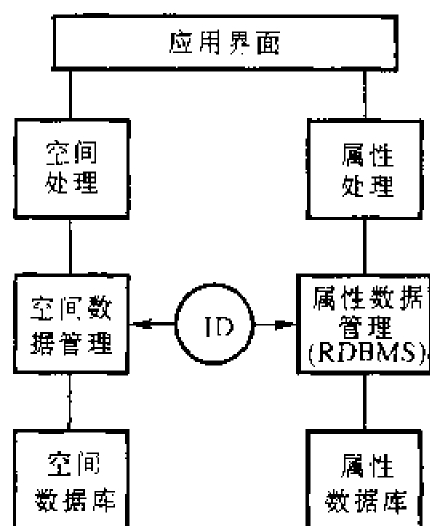


图 1－9 混合式数据库管理系统

2．扩展式

扩展式数据库管理系统是在标准 RDBMS 的顶层，通过将地理结构查询语言（Geo SQL）转化成标准的 SQL 查询，借助索引数据的辅助关系实施空间索引操作（图 1－10）。属于这种管理方式的 GIS 软件有 System 9、SmallWorld、Geovision 等。

3．开放式

开放式数据库管理系统，是利用专门开发的 DBMS 来统一管理空间数据和属性数据（图 1－11）。属于这种管理方式的 GIS 软件有 Tigris、GEO++、

GeoTropics 等。

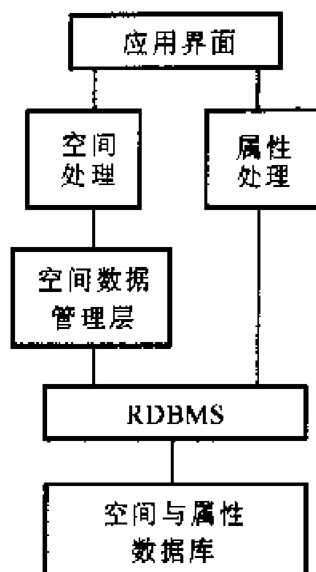


图 1-10 扩展式数据库管理系统

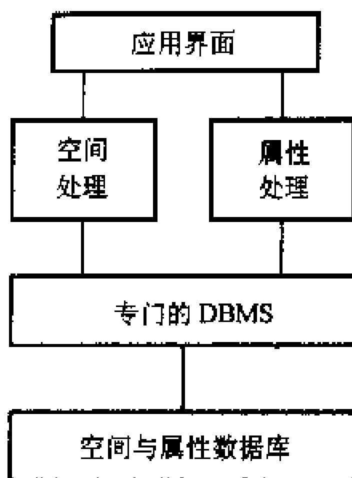


图 1-11 开放式数据库管理系统

空间数据库是 GIS 的重要组成和应用资源，它的建立和维护是一项非常复杂的工作，其技术也在不断完善中，其中空间数据库引擎(SDE)代表着这一技术的最新进展。

2.4 应用人员

GIS 应用人员包括系统开发人员和 GIS 技术的最终用户，他们的业务素质和专业知识是 GIS 工程及其应用成败的关键(图 1-12)。

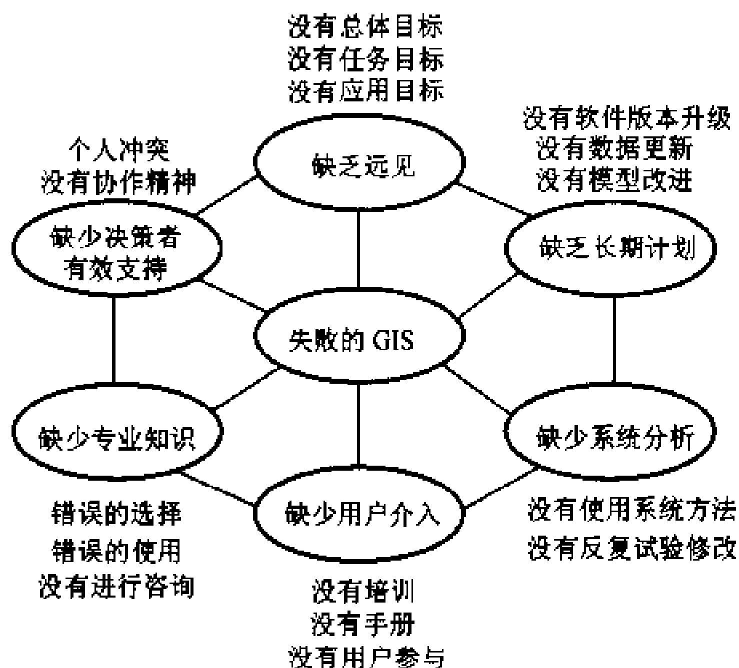


图 1-12 地理信息系统开发成功的要素构成图

(据陈述彭等，1999)

GIS的开发是一项以人为本的系统工程,包括用户机构的状况分析和调查、机构GIS系统开发目标的确定、系统开发的可行性分析、系统开发方案的选择和总体设计书的撰写等。开发人员要特别重视对用户机构的状况和要求进行具体分析,切忌只注重技术的细节。

系统开发过程中,对具体开发策略的确定、系统软件硬件的选型和空间数据库的建立等,系统开发人员必须根据GIS工程建设的特点和要求,并在深入调查研究的基础上,使确定的开发策略能适应GIS用户随时变化的需求,使系统的软硬件投入获得较高的效益回报,以及使建立的数据库能具有完善的质量保证。

在使用GIS时,应用人员不仅需要对GIS技术和功能有足够的了解,而且需要具备有效、全面和可行的组织管理能力。尤其在当前GIS技术发展十分迅速,为使现行系统始终处于优化的运作,其组织管理和维护的任务包括:GIS技术和管理人员的技术培训、硬件设备的维护和更新、软件功能扩充和升级、操作系统升级、数据更新、文档管理、系统版本管理、数据共享性建设等。

2.5 应用模型

GIS应用模型的构建和选择也是系统应用成败至关重要的因素,虽然GIS为解决各种现实问题提供了有效的基本工具,但对于某一专门应用目的的解决,必须通过构建专门的应用模型,例如土地利用适宜性模型、公园选址模型、洪水预测模型、人口扩散模型、森林增长模型、水土流失模型、最优化模型、影响模型等。

这些应用模型是客观世界中相应系统经由观念世界到信息世界的映射,反映了人类对客观世界利用改造的能动作用,并且是GIS技术产生社会经济效益的关键所在,也是GIS生命力的重要保证,因此在GIS技术中占有十分重要的地位。

构建GIS应用模型,首先必须明确用GIS求解问题的基本流程(图1-13);其次根据模型的研究对象和应用目的,确定模型的类别、相关的变量、参数和算法,构建模型逻辑结构框图;然后确定GIS空间操作项目和空间分析方法;最后是模型运行结果验证、修改和输出。显然,应用模型是GIS与相关专业连接的纽带,它的建立绝非是纯数学技术性问题,而必须以坚实而广泛的专业知识和经验为基础,对相关问题的机理和过程进行深入的研究,并从各种因素中找出其因果关系和内在规律,有时还需要采用从定性到定量的综合集成法,这样才能构建出真正有效的GIS应用模型。

这样,大量应用模型的研究、开发和应用,凝聚和验证了许多专家的经验和知识,无疑也为GIS应用系统向专家系统的发展打下基础。

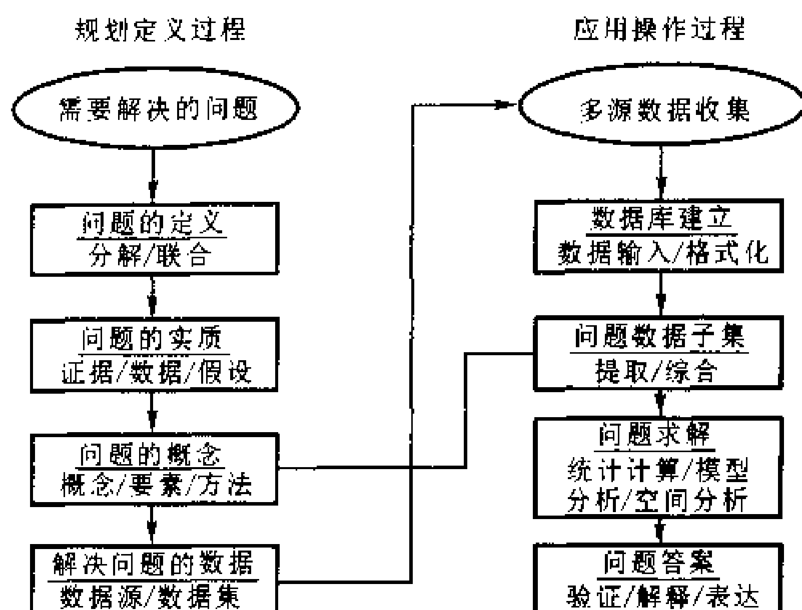


图 1-13 用 GIS 求解问题的基本流程图
(据陈述彭等,1999)

§3 GIS 的功能简介

由计算机技术与空间数据相结合而产生的 GIS 这一高新技术,它包含了处理地理信息的各种高级功能,但是它的基本功能是数据的采集、管理、处理、分析和输出。GIS 依托这些基本功能,通过利用空间分析技术、模型分析技术、网络技术、数据库和数据集成技术、二次开发环境等,演绎出丰富多彩的系统应用功能,满足社会和用户的广泛需求。

3.1 基本功能

1. 数据采集与编辑

地理信息系统的数据通常抽象为不同的专题或层(图 1-14)。数据采集与编辑功能就是保证各层实体的地物要素按顺序转化为 x、y 坐标及对应的代码输入到计算机中,各类数据的转化和输入方法如图 1-15 所示。

2. 数据存储与管理

数据库是数据存储与管理的最新技术,是一种先进的软件工程。GIS 数据库是区域内一定地理要素特征以一定的组织方式存储在一起的相关数据的集合。由于 GIS 数据库具有数据量大,空间数据与属性数据具有不可分割的联系,以及空间数据之间具有显著的拓扑结构等特点,因此 GIS 数据库管理功能,除了与属性数据有关的 DBMS 功能之外,对空间数据的管理技术主要包括:空间数据

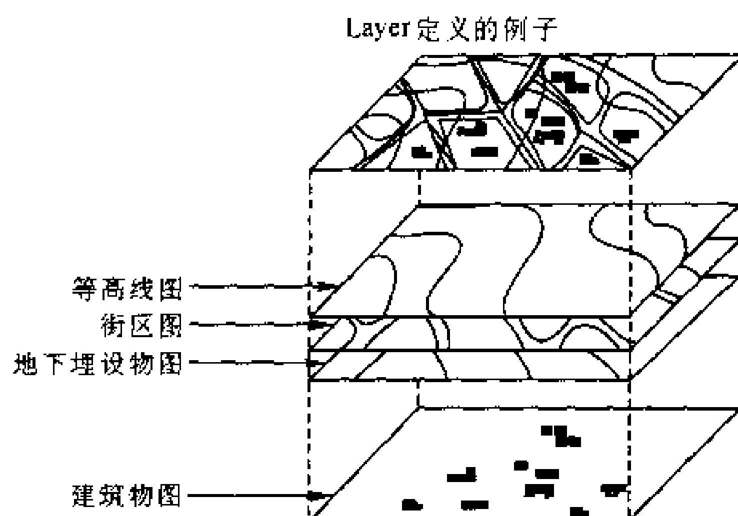


图1-14 GIS数据分层概念

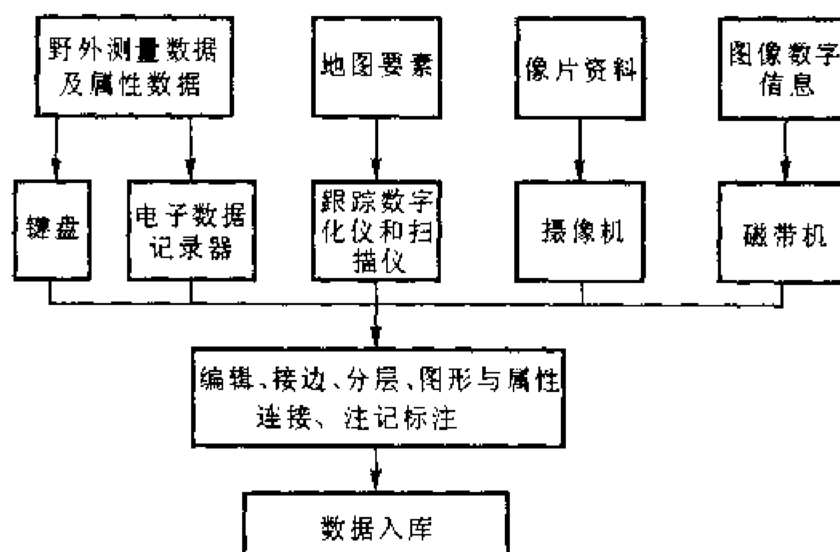


图1-15 GIS数据采集流程图

库的定义、数据访问和提取、从空间位置检索空间物体及其属性、从属性条件检索空间物体及其位置、开窗和接边操作、数据更新和维护等。

3. 数据处理和变换

由于GIS涉及的数据类型多种多样，同一种类型数据的质量也可能有很大的差异。为了保证系统数据的规范和统一，建立满足用户需求的数据文件，数据处理是GIS的基础功能之一。数据处理的任务和操作内容有：

(1) 数据变换。指对数据从一种数学状态转换为另一种数学状态，包括投影变换、辐射纠正、比例尺缩放、误差改正和处理等；

(2) 数据重构。指对数据从一种几何形态转换为另一种几何形态，包括数据拼接、数据截取、数据压缩、结构转换等；

(3) 数据抽取。指对数据从全集合到子集的条件提取,包括类型选择、窗口提取、布尔提取和空间内插等。

4. 空间分析和统计

空间分析和统计功能是GIS的一个独立研究领域,它的主要特点是帮助确定地理要素之间新的空间关系,它已经不仅成为区别于其他类型系统的一个重要标志,而且为用户提供了灵活地解决各类专门问题的有效工具。

(1) 拓扑叠合。通过将同一地区两个不同图层的特征相叠合,不仅建立新的空间特征,而且能将输入的特征属性予以合并,易于进行多条件的查询检索、地图裁剪、地图更新和应用模型分析等。

(2) 缓冲区建立。它是研究根据数据库的点、线、面实体,自动建立各种类型要素的缓冲多边形,用以确定不同地理要素的空间接近度或邻近性。它是GIS重要的和基本的空间分析功能之一,例如规划建设一个开发区,需要通知一定范围内的居民动迁;在林业规划中,需要按照距河流一定纵深范围来确定森林砍伐区,以防止周边地区的水土流失等等。

(3) 数字地形分析。GIS提供了构造数字高程模型及有关地形分析的功能模块,包括坡度、坡向、地表粗糙度、山谷线、山脊线、日照强度、库容量、表面积、立体图、剖面图和通视分析等,为地学研究、工程设计和辅助决策提供重要的基础性数据。

(4) 空间集合分析。空间集合分析是按照两个逻辑子集给定的条件进行布尔逻辑运算。其算子及其运算结果如表1-1所示。

表1-1 算子及其运算结果

A	*B*	·*NOT*·*A*	*A*·*AND*·*B*	*A*·*OR*·*B*	*A*·*XOR*·*B*	*A*·*NOT*·*B*	…
1	1	0	1	1	0	1	·
1	0	0	0	1	1	0	·
0	1	1	0	1	1	1	·
0	0	1	0	0	0	0	·

注:表中,1表示"真";0表示"假"。

5. 产品制作与显示

GIS产品是指经由系统处理和分析,产生具有新的概念和内容,可以直接输出供专业规划或决策人员使用的各种地图、图像、图表或文字说明,其中地图图形输出是GIS产品的主要表现形式,包括各种类型的符号图、动线图、点值图、晕线图、等值线图、立体图等。

一个运行的GIS,其产品制作与显示的功能包括:设置显示环境,定义制图环境,显示地图要素,定义字形符号,设置字符大小和颜色,标注图名和图例,以

及绘图文件编辑等。

6. 二次开发和编程

为使GIS技术广泛应用于各个领域,满足各种不同的应用需求,它必须具备的另一个重要基本功能是二次开发环境,包括提供专用语言的开发环境、用户可在自己的编程环境中调用GIS的命令和函数,或者系统将某些功能做成专门的控件供用户的开发语言(C,C++,VB,VC++,Delphi…)调用等。这样,用户可以非常方便地编制自己的菜单和程序,生成可视化的用户应用界面,完成GIS的以下各项应用功能的开发。

3.2 应用功能

1. 资源管理

资源的清查、管理和分析是GIS应用最广泛的领域,也是目前趋于成熟的主要应用领域,包括森林和矿产资源的管理、野生动植物的保护、土地资源潜力的评价和土地利用规划,以及水资源的时空分布特征研究等。系统的主要任务是将各种来源的数据和信息有机地汇集在一起,GIS软件能在一个连续无缝的方式下管理大型的地理数据库,这种功能强大的数据环境允许集成各种应用,最终用户通过GIS的客户端软件,可直接对数据库进行查询、显示、统计、制图,以及提供区域多种组合条件的资源分析,为资源的合理开发利用和规划决策提供依据。

2. 区域规划

城市与区域规划具有高度的综合性,涉及资源、环境、人口、交通、经济、教育、文化、金融等因素,但是要把这些信息进行筛选并转换成可用的形式并不容易,规划人员需要切实可行的、实时强的信息,而GIS能为规划人员提供功能强大的工具,例如规划人员利用GIS对交通流量、土地利用和人口数据进行分析,预测将来的道路等级;工程人员利用GIS将地质、水文和人文数据结合起来,进行路线和构造设计;GIS软件帮助政府部门完成总体规划、分区,现有土地利用、分区一致性,空地、开发区和设施位置等分析工作,是实现区域规划科学化和满足城市发展的重要保证。

3. 国土监测

GIS方法和多时相的遥感数据,可以有效地用于森林火灾的预测预报、洪水灾情监测和淹没损失估算、土地利用动态变化分析和环境质量的评估研究等。例如黄河三角洲地区的防洪减灾研究表明,在ARC/INFO地理信息系统支持下,通过建立大比例尺数字地形模型和获取有关的空间和属性数据,利用GIS的叠合操作和空间分析等功能,可以计算出若干个泄洪区域内的土地利用及其面积,比较不同泄洪区内房屋和财产损失等,可以确定泄洪区内人员撤退、财产转移和

救灾物资供应的最佳路线,保证以最快的速度有效应付突发事件的发生。

4. 辅助决策

GIS利用拥有的数据和互联网传输技术,已经实现了电子商贸的革命,满足企业决策多维性的需求。当前在全球协作的商业时代,90%以上的企业决策与地理数据有关,例如企业的分布、客货源、市场的地域规律、原料、运输、跨国生产、跨国销售等。利用GIS迅速有效管理空间数据,进行空间可视化分析,确定商业中心位置和潜在市场的分布,寻找商业地域规律,研究商机时空变化的趋势,不断为企业创造新的商机,GIS和互联网已成为最佳的决策支持系统和威力强大的商战武器。

§4　GIS的发展透视

4.1　发展概况

地理信息系统是20世纪60年代中期开始逐渐发展起来的一门新的技术。由于40年代和50年代计算机科学、地图学和航空摄影测量技术的发展,逐渐产生利用计算机汇总各种来源的数据,借助计算机处理和分析这些数据,最后通过计算机输出一系列结果,作为决策过程的有用信息,这就产生最早的地理信息系统的基本框架。长期以来,地学工作者一直要研究存在于具体时空框架中的地理要素,最初存储这些要素的共同介质是模拟地图,使用和分析空间数据时,是通过由这种模拟地图和读图人员组成的简易系统来完成的,偶尔也借助一些简单的量算工具。经验表明,采用这种简易的系统,提取少量数据比较容易,但是要提取大量的地理信息,或者研究存在于多种地图要素之间的复杂关系时,就很困难。例如,当地学研究人员要用手工方法在两组或两组以上的数据之间研究其可能的相互关系,而这些空间数据的坐标系统不一致,比例尺不统一时,是很困难的,甚至是不可能的。因此,当50年代末和60年代初,计算机获得广泛应用以后,很快就被应用于空间数据的存储和处理,使计算机成为地图信息存储和计算处理的装置,将很多地图转换为能被计算机利用的数字形式。计算机分析地图内容并提供信息,是从自然资源的管理和土地规划任务开始的,在这个基础上诞生了世界上第一个地理信息系统——加拿大地理信息系统(CGIS)。这时地理信息系统的特征是和计算机技术的发展水平联系在一起的,表现在计算机存储能力小,磁带存取速度慢。机助制图能力较强,地学分析功能比较简单,实现了手扶跟踪的数字化方法,可以完成地图数据的拓扑编辑,分幅数据的自动拼接,开创了格网单元的操作方法,发展了许多面向格网的系统。例如哈佛大学的SYMAP是最著名的一例,另外还有GRID、MLMIS等系统。所有这些处理空间数

据的主要技术，奠定了地理信息系统发展的基础。这一时期，地理信息系统发展的另一显著标志，是许多有关的组织和机构纷纷建立，例如1966年美国成立城市和区域信息系统协会（URISA），1969年又建立州信息系统全国协会（NASIS），国际地理联合会（IGU）于1968年设立了地理数据收集和处理委员会（CGDSP）。这些组织和机构的建立，对于传播地理信息系统的知识和发展地理信息系统的技术，起了重要的指导作用。

20世纪70年代，计算机发展到第三代，不仅内存容量大增，运算速度达到10^{-6}秒级，而且输入－输出设备比较齐全，推出了大容量直接存取设备——磁盘，为地理数据的录入、储存、检索、输出提供了强有力的手段，特别是人机对话和随机操作的应用，可以通过荧屏直接监视数字化的操作，而且制图分析的结果能很快看到，并进行实时的编辑。这时由于计算机技术及其在自然资源和环境数据处理中的应用，促使地理信息系统迅速发展。例如从1970年至1976年，美国地质调查所就建成50多个信息系统，分别作为处理地理、地质、水资源等领域空间信息的工具。其他如加拿大、前联邦德国、瑞典、日本等国也先后发展了自己的地理信息系统。地理信息系统的发展，使一些商业公司开始活跃起来，软件在市场上受到欢迎，同时管理问题也开始受到重视。例如，IGU较广泛深入地研究了5个主要系统的成功和失败，先后于1975年和1976年两次调查了与空间数据处理有关的计算机软件，以确定现有软件的类型、特点和质量，当时大约有320个程序组和系统，其中较完整的地理信息系统软件就有80个之多，并于1980年由美国地质调查所出版了《空间数据处理计算机软件》三卷一套的报告，基本总结了1979年以前世界各国地理信息系统发展的概貌。与此同时，Marble DF等拟订了处理空间数据的计算机软件说明的标准格式，对全部软件进行了系统的分类，提出地理信息系统的今后发展应着重研究空间数据处理的算法、数据结构和数据库管理系统三个方面。先后召开了一系列地理信息系统的国际讨论会，例如，1971年在法国Saint-Maximin召开了关于数据库的国际专家会议，国际地理联合会先后于1970年和1972年两次召开关于地理信息系统的学术讨论会，1978年FIG规定第三委员会的主要任务是研究地理信息系统，同年在前联邦德国的达姆斯塔特工业大学召开了第一次地理信息系统讨论会等。这期间，许多大学（例如美国纽约州立大学布法罗校区等）开始注意培养地理信息系统方面的人才，创建了地理信息系统实验室。因此，地理信息系统这一技术受到了政府部门、商业公司和大学的普遍重视，成为一个引人注目的领域。

20世纪80年代是地理信息系统普遍发展和推广应用的阶段。由于大规模和超大规模集成电路的问世，推出了第四代计算机，特别是微型计算机和远程通讯传输设备的出现，例如1984年Prime 9950象征着新一代微机的诞生，有4～6兆内存，每秒400万次，为计算机的普及应用创造了条件，加上计算机网络的建

立,使地理信息的传输时效得到极大的提高。在系统软件方面,完全面向数据管理的数据库管理系统(DBMS)通过操作系统(OS)管理数据,系统软件工具和应用软件工具得到研制,先后推出了 ARC/INFO、GENAMAP、MICROSTATION 和 SYSTEM9 等 GIS 基础软件。数据处理开始和数学模型、模拟等决策工具结合。地理信息系统的应用从解决基础设施的规划(如道路、输电线等)转向更加复杂的区域开发,例如土地的农业利用、城市化的发展、人口的规划和安置等,地理因素成为投资标准决策不可缺少的依据。这时期,地理信息系统不仅引起工业化国家的普遍兴趣,例如 1987 年美国成立了国家地理信息与分析中心(NCGIA),英、法、前联邦德国、挪威、瑞典、荷兰、以色列、澳大利亚、前苏联等国都在积极解决地理信息系统的发展和应用,而且不再受国家界线的限制,地理信息系统开始用于解决全球性的问题,例如全球的沙漠化、全球可居住区的评价、厄尔尼诺现象及酸雨、核扩散等对世界环境潜在的影响等。这时期,我国 GIS 的基础研究和应用研究也取得突破性的进展,例如成立了地理信息系统相关的国家重点实验室,开展了 GIS 的国家科技攻关项目研究等。因此,国际著名的地理信息系统专家 R.F.汤姆林逊认为,如果 20 世纪 70 年代是这个领域发展的巩固时期,那么 80 年代是国际地理信息系统发展具有突破性的年代。

20 世纪 90 年代是 GIS 应用的大发展,GIS 成为一种通用的地理信息技术工具被广泛应用,称为 GIS 的用户时代。一方面,许多机构逐渐了解 GIS 的功能,利用 GIS 作为必备的工作系统,改变传统的工作模式,提高工作效率和质量。另一方面,社会对 GIS 认识普遍增强,用户需求迅速增加,导致 GIS 应用领域的扩大和应用水平的提高。国家级乃至全球性的 GIS 系统还特别受到政府部门的关注,例如地理信息系统已列入美国政府制定的“信息高速公路”计划;美国副总统戈尔提出的“数字地球”战略,我国的“21 世纪议程”和“三金工程”等,都包含有地理信息系统项目的建设。

当前,GIS 正向着集成化、产业化和社会化发展方向迈进,呈现以下主要发展态势:

1. GIS 已成为一门综合性技术

近年来,随着 GIS 技术的发展和社会需求的增大,GIS 已经和计算机、通信等技术一样,成为信息技术(IT)的重要组成部分,GIS 需要应用其他信息技术,其他信息技术也需要应用 GIS。GIS 不但与全球定位系统(GPS)和遥感(RS)相结合,构成“3S”集成系统,而且与 CAD、多媒体、通信、因特网、办公自动化、虚拟现实等多种技术相结合,构成了综合的信息技术。

GIS 成为一门综合性技术的另一个显著标志是其自身系统功能和支持功能的不断完善,体现在:GIS 信息规范共享措施的建立;多源空间数据无缝集成和融合技术的实施;GIS 图形和属性数据的无缝结合与管理;大型 GIS 地理数据库

的无缝组织;空间数据质量控制和数据交换格式方案的解决;空间数据仓库和元数据技术的应用;国产 GIS 基础软件和组件式软件的开发与推出;多维和动态数据模型的应用设计;GIS 空间分析模型及体系的建立;开放式 GIS 技术的构建;以及 GIS 体系结构和解决方案的不断优化组合等等。因此,GIS 已经能够为用户提供地球表层及其附近的空间和非空间数据的获取、处理、分析、表示和传输的重要技术手段,这些技术的不断发展为数字地球的建立及其应用(图 1-16)提供了可靠的技术保证,也为人类造福提供了可靠的技术保证。

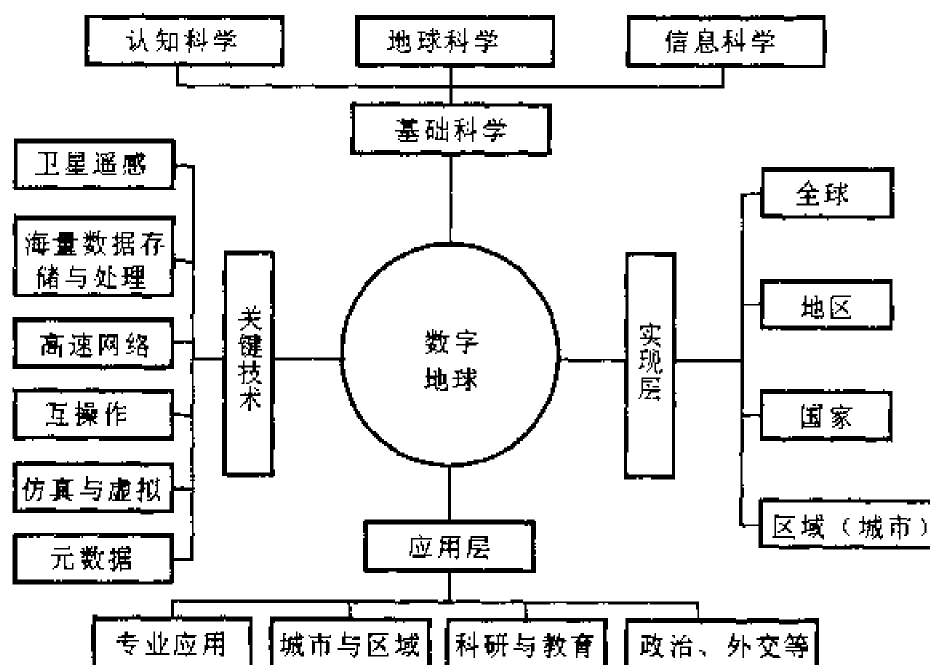


图 1-16 数字地球的框架

2. GIS 产业化的发展势头强劲

由于 GIS 是一项面向 21 世纪信息时代,关系国家综合竞争实力的高新技术,因此 GIS 及其产业化的发展日益受到各国的普遍关注。以美国为例,由 1:1 000 000至 1:24 000 的基础地理信息数字化产品和大批专业地学数字化信息产品已进入市场,以相当低廉的价格提供用户使用,约 85%的联邦机构和地方政府机构使用 GIS,每年 GIS 的应用项目达 1 万多个。90 年代全球 GIS 产业以每年 15%~40%的速度增长,预计 2000 年其产值将至少达 500 亿美元,成为信息产业中市场前景十分广阔,又相对独立的新兴产业。根据国际权威信息评估机构的划分,GIS 产业对象包括:硬件、软件、数据采集与转换、电子数据、遥感数据获取与处理、系统开发与集成和咨询与技术服务七个方面。其中,GIS 软件的发展经历了从早期的功能处理模块,发展到组件式 GIS 和 WebGIS 的过程。组件式 GIS 和 WebGIS 已经成为许多大型 GIS 公司产品的开发方向和竞争对象,相继推出了各自的产品(表 1-2)。

我国GIS产业发展也已被确定为国家信息产业中重要的发展方向，先后建成1:100万地形数据库和地名数据库、1:50万数字地理底图数据库、1:25万地形数据库、数字高程模型库。1:25万地形数据库含819幅图，包括水系、交通、境界、居民点、地形、植被等14层要素。数字高程模型库分为100 m×100 m格网和3 s×3 s格网两种。1:25万地名数据库共有805 431个地名。已推出ARC－China 1:100万光盘产品和1:400万中国数字产品等。目前正在加快建设1:5万基础地理数据库和全国七大江河重点防范区的1:1万空间数据库，为国民经济信息化建设提供坚实的基础。

3. GIS网络化已构成当今社会的热点

近几年来，因特网在全球的迅速发展，为信息产业提供了一次发展机会，也是GIS技术发展的一次良机。因特网是全球最大的、开放的、由众多网络互联而成的计算机互联网，它的分布式结构和TCP/IP协议的互联灵活性是其成功的关键，而其中全球信息网(Web)服务是在因特网基础上发展起来的，它的发展为地理信息共享提供了一个开放的信息空间。Web以HTTP(超文本传输协议)为信息通信协议，采用B/S(浏览器/服务器)的结构，这种结构解决了传统的C/S(客户机/服务器)结构的客户端多平台问题，通过一个浏览器可以访问多个应用服务器。同时，这种结构将C/S计算体系的两层结构自然延伸为三层甚至多层结构(图1－17)。这种基于www的地理信息共享体系结构，以浏览器为前端，采用标准对象开发平台，例如JAVA、C++等开发可嵌入浏览器的应用组件，实现数据集的检索和对服务器端的数据进行浏览。后端由www服务器、应用服务器、数据库服务器和搜索引擎组成。用户通过元数据的查询和搜索引擎，可以获取各种地理空间数据和属性数据，进行地理空间的分析、预测、推理、决策等，以及用户在查找到感兴趣的数据后，可以通过FTP等不同的方式进行下载，极大地拓展了GIS的功能。

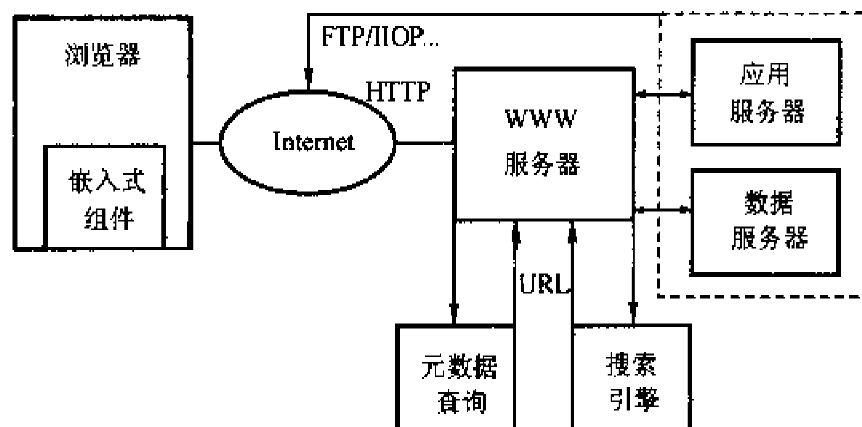


图1－17 基于www的信息共享体系结构

由于WebGIS具有廉价、实用和易推广等特点，目前各大GIS软件厂商纷纷推出各自的WebGIS解决方案，如表1-2所示，这些WebGIS软件有ESRI公司的IMS、Intergraph公司的Geomedia Web Map和Geomedia Web-Enterprise、MapInfo公司的Mapxtreme、AutoDesk公司的MapGuide等。它们的开发和应用，使传统GIS面临着全新的发展，将使GIS迅速走向全社会，即GIS的社会化正向我们走来！

表1-2 许多大型GIS公司开发的组件式GIS和Web GIS

研制单位	组件式GIS	WebGIS
ESRI	MapObjects	Internet Map Server
Intergraph	Geomedia	Geomedia Web Map
MapInfo	Mapx	Mapxtreme
Autodesk		MapGuide
Bently		Model Server/Dicovery
长地公司		MapGIS Web

4. 地理信息科学的产生和发展

目前在全球范围内，GIS以前所未有的发展速度在科技界和企业界推广应用，GIS的推广应用大大提高了人类处理和分析有关地球资源、环境、社会与经济数据的能力，而GIS技术及其应用的进一步发展必须依托基础理论的研究，发展地理信息科学。

法国大地测量和摄影测量学家Bernart Dubuisson于1975年首先将地球量测(法文名Geomatique)正式用于科学文献。1990年Gagnon将"Geomatics"定义为"利用各种手段，通过一切途径来获取和管理有关空间基础信息的空间数据部分的科学技术领域"。随即，加拿大、澳大利亚、英国、荷兰、香港等国家和地区的一些高等学校相关系科及相关的政府机构等出现了更名热潮。例如，加拿大拉瓦尔大学、卡尔加里大学、新不伦瑞克大学将测量工程系改名为"Geomatics"系，荷兰ITC成立了"Geoinfomatics"系，加拿大能源矿产资源部将其测绘和遥感局改名为"Geomatics Canada"。从上述"Geomatics"一词出现的过程可以看出，"Geomatics"反映了现代测绘科学、遥感和地理信息系统与现代计算机科学和信息科学相结合的多学科集成，以满足对空间信息要求的发展趋势。

王之卓教授从学科发展的高度提出使用图像信息学(Iconic Informatics)来概括与测绘有关的学科，图像信息学主要是处理影像数据，而地理信息学则是处理基于地学参照系的数据。地理信息学的研究内容正如陈述彭教授所指出的，它不仅仅限于研究物质流与能量流的信息载体，而且包括研究地学信息流程的动力学机理与时空特征、地学信息传输机理及其不确定性与可预见性等，并认为

Geoinformatics 不同于 Geomatics，在于这个 Info 还包括很多地学规律，其分析模型必须以地学为基础。

Goodchild 于 1992 年提出地理信息科学（Geographic Information Science）的概念，他认为地理信息科学主要研究在应用计算机技术对地理信息进行处理、存储、提取以及管理和分析过程中提出的一系列基本问题，如数据的获取和集成、分布式计算、地理信息的认知和表达、空间分析、地理信息基础设施建设、地理数据的不确定性及其对于地理信息系统操作的影响、地理信息系统的社会实践等。

地理信息科学的提出是 GIS 技术及其应用发展到一个相当水平后的必然要求，但作为一门新兴发展的分支学科，其科学理论与方法还处于初期发展阶段，学科体系尚未健全，但从学科的基本性质看，它由三部分组成：地球信息机理是其理论研究的主体，地理信息技术是其研究手段，全球变化与区域可持续发展是其主要应用研究领域（图 1－18）。

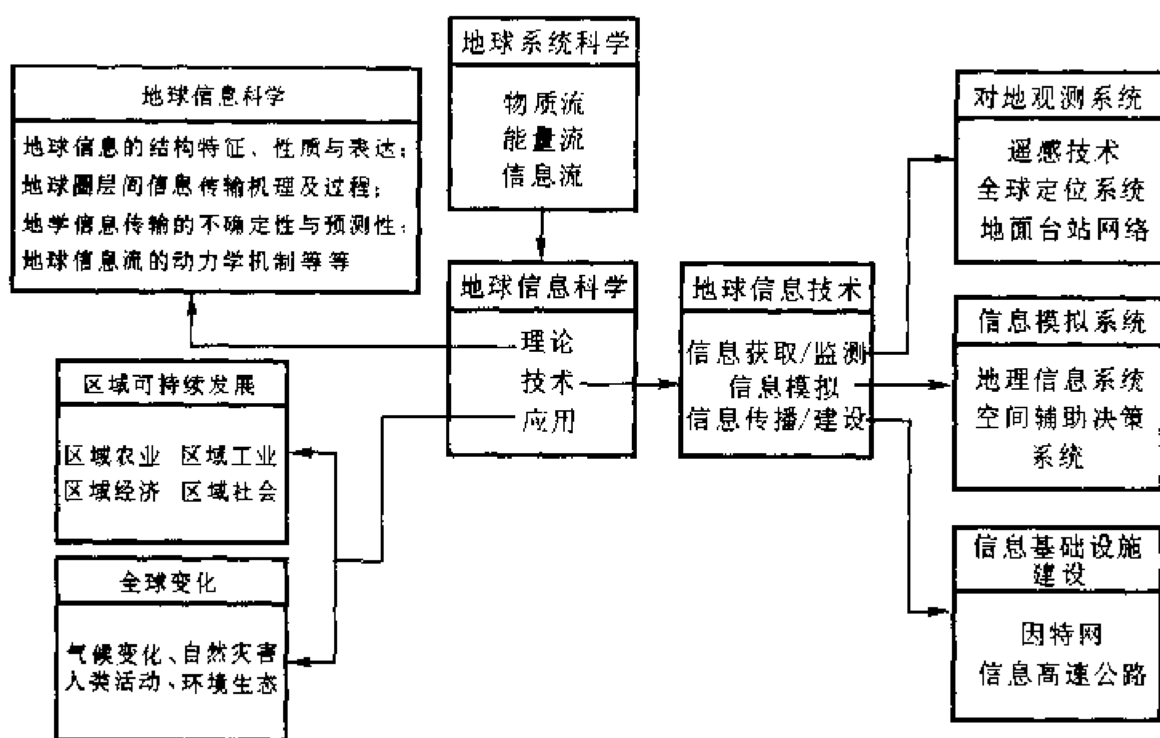


图 1－18　地理信息科学的科学体系
（据陈述彭等，2000）

4.2　基础理论

地理信息系统是传统科学与现代技术相结合而诞生的边缘学科，因此它明显地体现出多学科交叉的特征，这些交叉学科的基础理论同样构成地理信息系统的基础理论体系。这些学科包括地理学、地图学、测量学、数学、统计学、计算

机科学以及一切与获取、处理和分析空间数据有关的学科与技术。

在GIS的相关学科中，首先是地理学。地理学是以地域为单元研究人类环境的结构、功能、演化以及人地的相互关系，它广泛涉及人类居住的地球和世界，这与GIS的研究对象是一致的。地理学中的空间分析历史悠久，而空间分析正是GIS的核心，地理学作为GIS的理论依托，为GIS提供引导空间分析的方法和观点。因此，美国学者Parker把地理学称为GIS之父，这是不为过的。

测绘学及其分支学科，如大地测量学、摄影测量学、地图学等，不但为GIS提供高精度的空间数据，而且它们中的误差理论，地图投影理论、图形学理论、许多相关的算法等，可直接用于GIS空间数据的处理，保证空间数据的精度和质量，以及GIS产品的开发和输出等。

GIS是地理空间数据与计算机技术相结合的产物，例如计算机科学的信息论，为GIS数据的组织、编目、定位、存储、检索和维护等，提供了信息模型、信息结构和数据管理的方法论依据，以及使得各种形式的空间数据能够在计算机中表示。数据库管理系统提供了各种用户共享而具有最小的冗余度和较高的数据与应用程序的独立性、安全性和完整性的机制保证。

数学的许多分支学科，包括几何学、统计学、运筹学、数学形态学、图形代数、拓扑学、图论、分形分维理论等，已经广泛地应用于GIS空间数据的分析，例如路径和网络分析、数字地形模型分析、缓冲带分析和许多应用模型的构建等。

空间实体的表达和空间关系描述是GIS中最基础的科学问题。自然界中的空间实体分为规则和不规则两种类型，它们都是采用空间数据来表达。空间数据又分为矢量和栅格两种形式，根据矢量图形复杂程度的不同，分别应用欧几里德(Euclid)几何学中的点、线、面及其组合体和分形几何学(Fractal Geometry)中的分数维来表达和描述它们，初步解决了规则空间实体和不规则空间实体的不同表达方法，这是数学对GIS基础理论作出的贡献，而且随着随机几何学(Stochastic Geometry)的应用，还将有可能形成一套全新的空间实体表达理论。

美国大学地理信息科学协会(UCGIS)，于1995年发表的地理信息科学的优先研究领域和美国国家地理信息与分析中心(NCGIA)，于1996年发表的高级地理信息科学中，都提到了地理信息的认知理论等问题。地理信息的认知和地理信息机理是地理信息科学的重要基础理论之一。认知的基本原理是：所有的物体都有自己特定的物理和能量特征。物理特征如形状、大小、结构等外部特征，以及组成成分和结构等内部特征。能量特征包括重力、磁力、声波及电磁波谱(γ、X射线、可见光、红外到无线电波)等特征。凡是属于同一属性的物质具有相同的物理特征和能量特征，也就是根据物理的和能量的特征差异来识别物体的类别。因此，认知论和认知图式(图1-19)对于复杂地理空间和地理过程的分析、GIS的信息转换和GIS图形用户界面的设计与开发等，都具有重要的指导

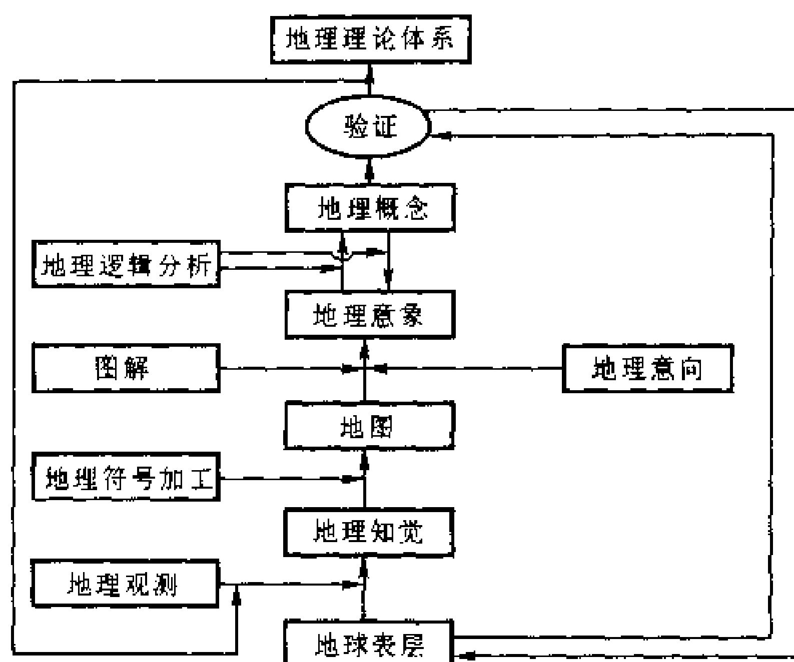


图1-19 地理学认知图示
(据鲁学军,1996)

意义。

近年倡导和兴起的资源环境信息图谱的研究,是这种认知理论研究的重要组成部分。图谱的图是指图形,谱是指系列,图谱是指图形系列。图谱分基本图谱(如地形地貌图谱、森林图谱等)和综合图谱(如征兆图谱、诊断图谱)。根据征兆图谱可以预测天气的变化、作物的长势等;根据诊断图谱可以分析某地区受灾状况、诊断某地区城市化的速度等。这种信息图谱的研究将可能成为GIS的基础理论之一。

李德仁教授在《关于地理信息理论的若干思考》一文中,提出了十二个问题需要我们从理论上去加以解决,并就如何建立地理信息理论,提出三个需要注意的观念上的问题,这就是:要充分认识到计算机技术进步带来的机遇,要打破传统地图数据库对GIS发展形成的框框,要充分发挥现代数学理论和工具在地理信息理论研究中的作用。这些问题和意见,对于呼唤地理信息理论的建立具有重要的启示和催化意义。

美国加州大学Goodchild MF教授在《地理信息科学发展》一文中指出:GIS的技术导引(Technology-Driven)和应用导引(Application-Driven)两大特点丝毫不影响GIS的研究特性和逐步形成它作为一门科学的作用。为保证GIS技术导引的功能,已经形成GIS的技术基础理论体系,包括遥感技术、定位技术、通讯技术(光纤通讯、移动电话、电视、卫星通讯、数字通讯)、计算机技术(数值计算、信息处理、自动控制、专家系统、数据库和数据仓库技术、多媒体技术、网络技术、软件

构件化技术、虚拟现实技术等)。为实现这些技术的有机集成,必须研究遥感成像机理、遥感信息机理、遥感数据的反演方法、模拟数据与数字数据的自动转换、定位原理、数据模型和时空数据结构、多媒体和超媒体的表示、一体化管理及系统集成的理论和方法等。这些理论和方法的解决,新一代 GIS 也才有可能实现和建立。

当今由于信息技术的不断发展,信息已成为国民经济中的重要支柱产业之一,使社会逐步发展到信息经济或知识经济时代。地理信息的应用,如数字农业、数字城市、数字区域和数字地球等,将形成大的产业,推动社会经济的发展,这时地理信息作为一种商品,它在市场上的流通,必须遵循商品的基本规律,因此要研究地理信息经济学、地理信息法律学和地理信息符号学。实际上,UCGIS 于 1995 年提出的 GIS 优先研究领域中,已经包含 GIS 与社会这个课题。

在全球范围内,科学家们曾开展了"GIS:是工具还是科学?"的大讨论,多数科学家认为地理信息的研究和方法应该是科学。因为它已经具备了成为科学的条件,如有明确的研究对象和研究手段,有自己的理论体系,有一定数量的研究队伍,有明确的应用领域和广泛的服务对象,因此已经由技术发展成为科学,即 Geoinformatics。

思考题

1. 什么是地理信息系统(GIS)? 它与一般的计算机应用系统有哪些异同点?

2. GIS 由哪几个主要部分组成? 它的基本功能有哪些? 试以目前广泛应用的两个基础 GIS 软件为例,列出它们的功能分类表,并按重要性进行排队和比较其异同点。

3. 根据你的了解,阐述 GIS 的相关学科及关联技术,并就 GIS 基础理论的建立和发展问题,发表你的意见和观点。

4. GIS 可应用于哪些领域? 试结合你的专业论述 GIS 的应用和发展前景。

5. 你对 GIS 社会化的发展趋势是怎么理解的?

6. 给出"Geomatics"一词的定义。如何理解"Geo"和"matics"的含义? "Geomatics"的出现有什么意义?

第2章 地理信息系统的数据结构

§1 地理空间及其表达

1.1 地理空间的概念

地理空间上至大气电离层，下至地幔莫霍面，是生命过程活跃的场所，也是宇宙过程对地球影响最大的区域。

地理信息系统中的空间概念常用“地理空间”(geo-spatial)来表述，一般包括地理空间定位框架及其所联结的特征实体。地理空间定位框架即大地测量控制，由平面控制网和高程控制网组成。大地测量控制为建立所有的地理数据的坐标位置提供了一个通用参考系，利用该通用参考系可以将全国范围使用的平面及高程坐标系与所有的地理要素相连接。大地测量控制信息的主要要素就是大地测量控制点，这些设标点(有时为动态的 GPS 控制点)的平面位置和高程被精确地测量，并用于其他点位的确定。因此，大地测量控制信息在开发所有的框架数据及用户的应用数据中发挥着关键的作用。

目前，我国采用的大地坐标系为 1980 年中国国家大地坐标系，该坐标系选用 1975 年国际大地测量协会推荐的国际椭球(图 2-1)，其具体参数为：

赤道半径$(a)=6\,378\,140.000\,000\,000\,0$ m

极半径$(b)=6\,356\,755.288\,157\,528\,7$ m

地球扁率$(f)=\dfrac{a-b}{a}=\dfrac{1}{298.257}$

1980 年中国国家大地坐标系的大地原点，设在我国中部的陕西省泾阳县永乐镇，简称西安原点。由于大地原点选在我国中部位置，可以减少坐标传递误差的积累，而且该坐标系在定向上与我国目前使用的 1968.OJYD(地极原点)的方向相一致，起始大地子午面平行于格林尼治平均天文台子午面，以及在我国境内，该坐标系相应的地球椭球面和大地水准面最为密合，因此 1978 年决定建立 1980 年国家大地坐标系。

根据不同需求，我国现有三种大地坐标系并存：一是 1954 年北京坐标系(局

部平差）；二是 1980 年国家大地坐标系（整体平差）；三是地心坐标系。对应于每一个坐标系统点的坐标，可以用大地坐标形式表示，即用(L,B,H)表示；也可以用空间大地直角坐标形式表示，即用(x,y,z)表示。不同坐标系统的坐标，通过一定数学模型的转换参数，在一定的精度范围内可以互相转换，具体转换方法详见参考文献[8]。其中，1954 年北京坐标系和 1980 年国家大地坐标系中的点的坐标，更多的是将其投影至高斯－克吕格投影平面，以平面坐标 x、y 形式表示，用于控制地形测图。

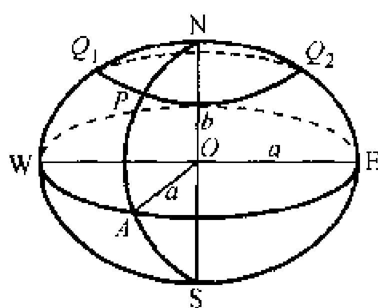


图 2－1 1975 年国际椭球元素及其关系示意图

在高斯－克吕格平面直角坐标中，x 表示纵轴，y 表示横轴。点的高斯－克吕格平面直角坐标是通过高斯－克吕格投影计算得到的，我国各个等级大地点成果表中所载的坐标即为高斯－克吕格平面直角坐标。

将椭球面上各点的大地坐标，按照一定的数学法则，变换为平面上相应点的平面直角坐标，通常称之为地图投影。这里所说的一定的数学法则，可以用下面两个方程式表示：

$$x = f_1(L,B)$$
$$y = f_2(L,B)$$

式中：(L,B)是椭球面上某一点的大地坐标；

(x,y)是该点投影在投影平面上的直角坐标。

地理信息系统中特征实体的位置，通常就是指经过投影变换后平面上的直角坐标。

高程指空间参考的高于或低于某基准平面的垂直位置，主要用来提供地形信息。我国现在规定的高程起算基准面为“1985 国家高程基准”，该基准比原国务院批准启用的“黄海平均海平面”高 29 mm。

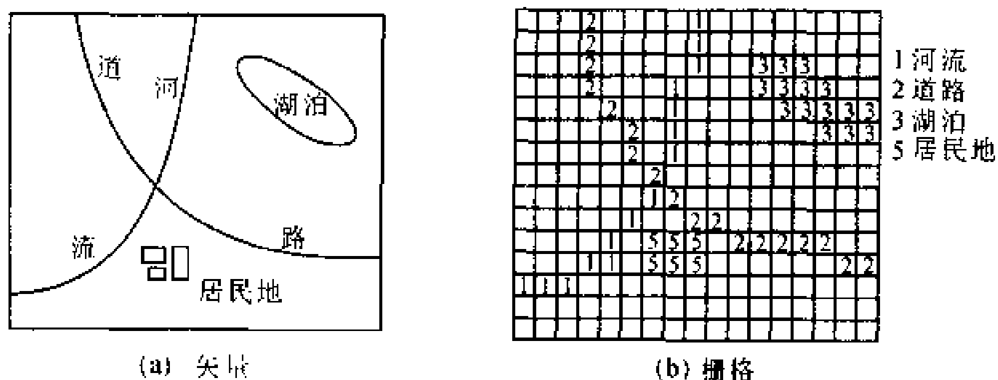
研究地理空间，除了建立地理空间的定位参考框架，还必须分析地理空间特征实体或地理空间信息的几何形态和时空分布规律及其相互之间的关系。地理空间特征实体指具有形状、属性和时序特征的空间对象或地理实体，包括点、线、面、曲面和体，它们构成地球圈层间复杂的地理综合体，也是地理信息系统表示和建库的主要对象。

1.2 空间实体的表达

如前所述，地理空间的特征实体包括点(point)、线(line)、面(polygon)、曲面(surface)和体(volume)等多种类型，如何以有效的形式表达它们，关系到计算机识别、存储、处理的可能性和有效性。

在计算机中，现实世界是以各种数字和字符形式来表达和记录的，基于计算机的地理信息系统不能直接识别和处理各种以图形形式表达的特征实体，要使计算机能识别和处理它们，必须对这些特征实体进行数据表达。

当对特征实体进行数据表达时，关键又看如何表达空间的一个点，因为点是构成地理空间特征实体的基本元素。如果采用一个没有大小的点(坐标)来表达基本点元素时，称为矢量表示法(图2-2(a))；如果采用一个有固定大小的点(面元)来表达基本点元素时，称为栅格表示法(图2-2(b))。它们分别对应矢量数据模型和栅格数据模型，代表着从信息世界观点对现实世界空间目标的两种不同的数据表达方法，其在功能、使用方法及应用对象上都有一定的差异，这在一定程度上反映出GIS表示现实世界的不同概念，也是人类悟性的产物。



(a) 矢量　　　　(b) 栅格

图2-2　表示空间数据的两种基本形式

§2　地理空间数据及其特征

2.1　GIS的空间数据

空间数据是GIS的核心，也有人称它是GIS的血液，因为GIS的操作对象是空间数据，因此设计和使用GIS的第一步工作就是根据系统的功能，获取所需要的空间数据，并创建空间数据库。

GIS中的数据来源和数据类型繁多，概括起来主要有以下几种类型：

(1) 地图数据。来源于各种类型的普通地图和专题地图，这些地图的内容丰富，图上实体间的空间关系直观，实体的类别或属性清晰，实测地形图还具有很高的精度。

(2) 影像数据。主要来源于卫星遥感和航空遥感，包括多平台、多层面、多种传感器、多时相、多光谱、多角度和多种分辨率的遥感影像数据，构成多源海量

	点	线	面
类型数据	点状要素	线状要素	面状要素
面域数据	区域中心	境界线	行政单元
网络数据	道路交点	街道	街区
样本数据	气象站	航线	样方分布区
曲面数据	高程点	等值线	概略等值区
文本数据	地点名称	线状要素名称	区域名称
符号数据	点状符号	线状符号	面状符号

图 2-3　空间数据的类型和表示方法

（据 Jack Dangermond，1984）

数据，也是 GIS 的最有效的数据源之一。

（3）地形数据。来源于地形等高线图的数字化，已建立的数字高程模型（DEM）和其他实测的地形数据等。

(4) 属性数据。来源于各类调查报告、实测数据、文献资料、解译信息等。

(5) 元数据。来源于由各类纯数据通过调查、推理、分析和总结得到的有关数据的数据，例如数据来源、数据权属、数据产生的时间、数据精度、数据分辨率、源数据比例尺、数据转换方法等。

空间数据根据表示对象的不同，又具体分为七种类型(图2-3)，它们各表示的具体内容如下：

(1) 类型数据。例如考古地点、道路线、土壤类型的分布等。

(2) 面域数据。例如随机多边形的中心点，行政区域界线、行政单元等。

(3) 网络数据。例如道路交点、街道、街区等。

(4) 样本数据。例如气象站、航线、野外样方分布区等。

(5) 曲面数据。例如高程点、等高线、等值区域等。

(6) 文本数据。例如地名、河流名称、区域名称等。

(7) 符号数据。例如点状符号、线状符号、面状符号(晕线)等。

所有这些不同类型的数据都可以分为点、线、面三种不同的图形，并可以分别采用 x、y 平面坐标，地理经纬度 λ、φ，或者格网法表示。

2.2　空间数据的基本特征

在地理信息系统中，由于空间数据代表着现实世界地理实体或现象在信息世界中的映射，因此它反映的特征同样应该包括自然界地理实体向人类传递的基本信息。如图2-4所示，设该图为一幅交通图，它传递的基本信息包括：

(1) 三条呈不同分布状态的交通线。一条近乎直线，一条呈"S"形，另一条为环状，表示它们在地球表面上呈不同分布状态的交通线，称为定位信息。

(2) 三条分别具有不同等级的交通线。近乎直线的为主干道，呈"S"形的为次干道，成环状的为支路，称为属性信息。

(3) 三条互相具有关联关系的交通线。主干道与次干道在结点 N_2 处相联接，主干道的结点 N_1 和 N_2 相邻接，结点 N_2 分别与三条路段 C_1、C_2 和 C_3 相关联等等，称为拓扑信息。

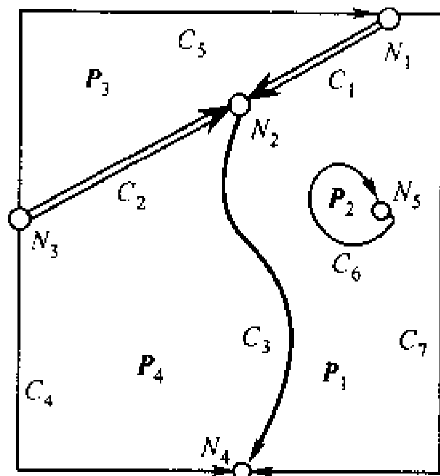


图2-4　地理实体：交通线

在GIS中，上述基本信息通过空间数据表达，分别对应着定位特征数据、属性特征数据和拓扑特征数据。对于随时间变化的地理实体或现象，还同时对应着时序或时间特征数据。一般地，空间特征数据包括地理实体或现象的定位数

据和拓扑数据，属性特征数据包括地理实体或现象的专题属性（名称、分类、数量等）数据和时间数据，而空间特征数据和属性特征数据统称为空间数据或地理数据。因此，空间数据的特征可以概括为空间特征和属性特征（图2-5）。

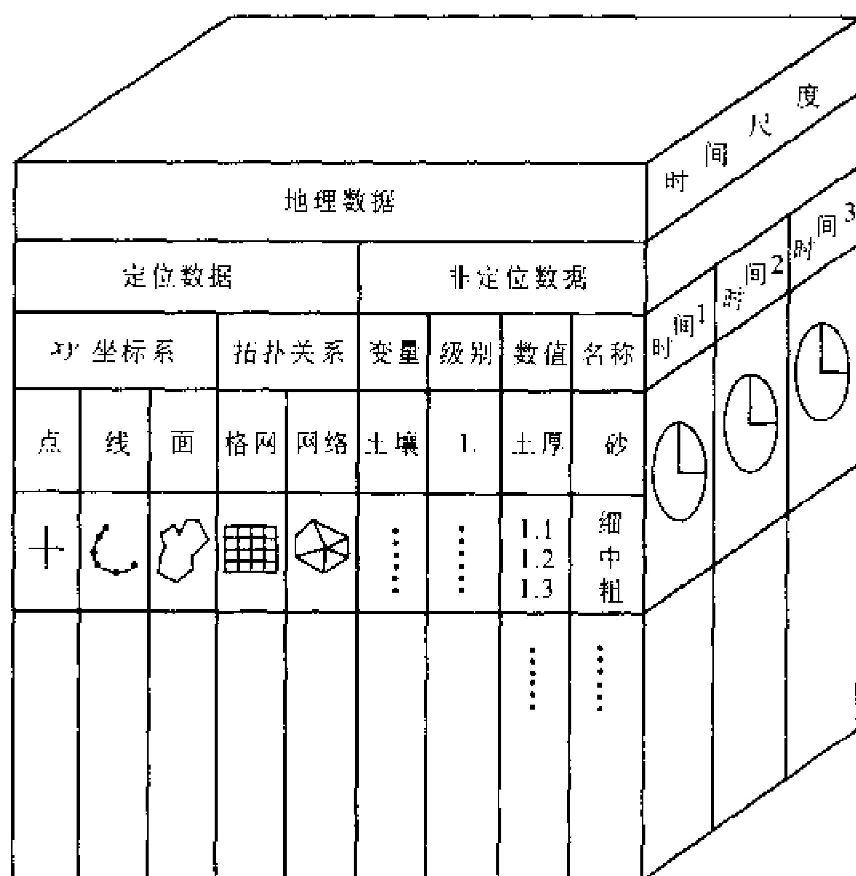


图2-5 空间数据的基本特征

（据 Jack Dangermond, 1984）

在地理信息系统中，对于凡具有网状结构特征的地理要素，例如自然与行政的分区、各种资源类型的空间分布以及交通网等，都存在结点、弧段和多边形之间的拓扑结构。拓扑结构是明确定义空间结构关系的一种数学方法，在地理信息系统中，它不但用于空间数据的编辑和组织，而且在空间分析和应用中都具有非常重要的意义。如图2-6所示，空间数据的拓扑关系包括：

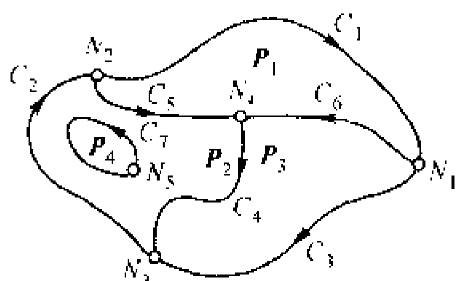
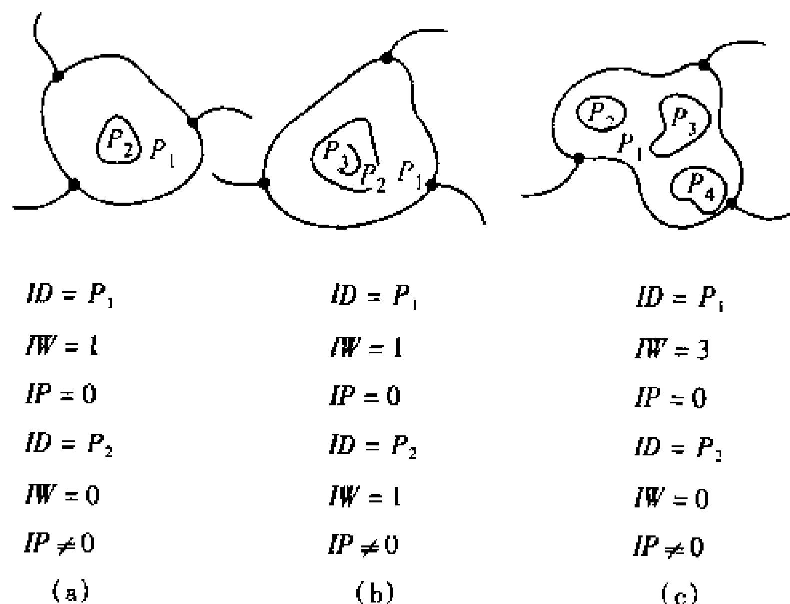


图2-6 空间数据的拓扑关系

(1) 拓扑邻接。指存在于空间图形的同类元素之间的拓扑关系。例如结点邻接关系 N_1/N_4，N_1/N_2，…；多边形邻接关系 P_1/P_3，P_2/P_3，…。

(2) 拓扑关联。指存在于空间图形的不同元素之间的拓扑关系，例如结点与弧段的关联关系 N_1/C_1、C_3、C_6；N_2/C_1、C_2、C_5，…；多边形与弧段的关联关系 P_1/C_1、C_5、C_6；P_2/C_2、C_4、C_5、C_7，…。

(3) 拓扑包含。指存在于空间图形的同类，但不同级的元素之间的拓扑关系。包含关系分简单包含、多层包含和等价包含三种形式。设 *ID* 表示当前多边形，*IW* 表示等价包含，*IP* 表示 *ID* 为岛（$IP>0$）或非岛（$IP=0$），则包含关系的形式如图 2-7 所示。



(a)	(b)	(c)
$ID=P_1$	$ID=P_1$	$ID=P_1$
$IW=1$	$IW=1$	$IW=3$
$IP=0$	$IP=0$	$IP=0$
$ID=P_2$	$ID=P_2$	$ID=P_2$
$IW=0$	$IW=1$	$IW=0$
$IP\neq 0$	$IP\neq 0$	$IP\neq 0$

图 2-7 拓扑包含关系的几种形式

如果要将结点、弧段和多边形之间的拓扑结构表达出来，可以形成四个关系表，如表 2-1、表 2-2、表 2-3 和表 2-4 所示。

表 2-1 结点与弧段的拓扑关系

结点	弧段
N_1	C_1, C_3, C_6
N_2	C_1, C_2, C_5
N_3	C_2, C_3, C_4
⋮	⋮

表 2-2 弧段与结点的拓扑关系

弧段	结点	
	from	to
C_1	N_2	N_1
C_2	N_3	N_2
C_3	N_1	N_3
⋮	⋮	⋮

表 2-3 弧段与多边形的拓扑关系

弧段	多边形	
	Left	Right
C_1	Φ	P_1
C_2	Φ	P_2
C_3	Φ	P_3
⋮	⋮	⋮

表 2-4 多边形与弧段的拓扑关系

多边形	弧段
P_1	$C_1, C_6, -C_5$
P_2	C_2, C_5, C_4, C_7
P_3	$C_3, -C_4, -C_6$
⋮	⋮

空间数据的拓扑关系，对地理信息系统的数据处理和空间分析，具有重要的意义，因为：

(1) 根据拓扑关系，不需要利用坐标或距离，可以确定一种地理实体相对于另一种地理实体的空间位置关系。因为拓扑数据已经清楚地反映出地理实体之间的逻辑结构关系，而且这种拓扑数据较之几何数据有更大的稳定性，即它不随地图投影而变化。

(2) 利用拓扑数据有利于空间要素的查询。例如应答像某区域与哪些区域邻接；某条河流能为哪些政区的居民提供水源；与某一湖泊邻接的土地利用类型有哪些；特别是野生生物学家可能想确定一块与湖泊相邻的土地覆盖区，用于对生物栖息环境作出评价等等，都需要利用拓扑数据。

(3) 可以利用拓扑数据作为工具，重建地理实体。例如建立封闭多边形，实现道路的选取，进行最佳路径的计算等等。

2.3 空间数据的计算机表示

如前所述，表示地理实体的空间数据包含着空间特征和属性特征，对具有这些复杂特征的空间数据，如何组织和建立它们之间的联系，以便计算机存储和操作，这称为数据结构。数据结构是 GIS 的核心技术，一直是 GIS 领域的重要研究课题。

以 ARC/INFO 基于矢量数据模型的系统为例，为了将空间数据存入计算机、首先，从逻辑上将空间数据抽象为不同的专题或层(图 1-14)，如土地利用、地形、道路、居民区、土壤单元、森林分布等，一个专题层包含指定区域内地理要素的位置数据和属性数据。其次，将一个专题层的地理要素或实体分解为点、线或面状目标，其中地理实体相邻两个结点间的一个弧段是基本的存储目标，每个目标的数据由定位数据、属性数据和拓扑数据组成。由基本目标构成数据库的逻辑过程如图 2-8 所示，即具有相同的分类码的同类目标组成类型，一类或相近的若干类构成数据层，若干数据层构成图幅，全部数据组成数据库。最后，对目

标进行数字表示，其中对每个弧段或目标分配一个用户标识码(User-ID)，弧段的位置和形状由一系列 x，y 坐标定义，弧段的拓扑关系由始结点、终结点、左多边形和右多边形四个数据项组成，弧段的属性数据存储在相应的属性表中。每个弧段的空间特征和属性特征通过用户标识码进行联接(图 2-9)。

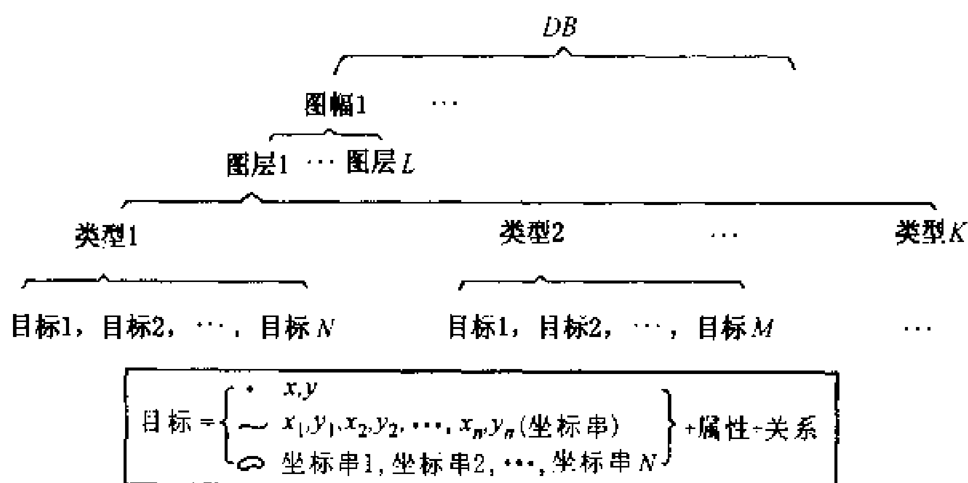
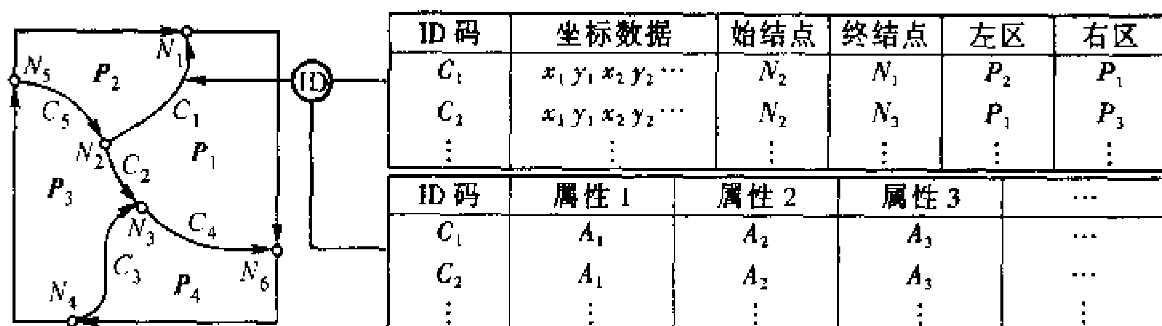


图 2-8 由目标构成数据库的逻辑过程



ID 码	坐标数据	始结点	终结点	左区	右区
C_1	$x_1y_1x_2y_2\cdots$	N_2	N_1	P_2	P_1
C_2	$x_1y_1x_2y_2\cdots$	N_2	N_3	P_1	P_3
⋮	⋮	⋮	⋮	⋮	⋮

ID 码	属性 1	属性 2	属性 3	…
C_1	A_1	A_2	A_3	…
C_2	A_1	A_2	A_3	…
⋮	⋮	⋮	⋮	⋮

图 2-9 一个空间目标的空间特征与属性特征之间的联系

空间数据的这种表示方法，是以地理实体相邻两个结点间的一个弧段作为基本存储目标，每个弧段都是一个有向弧段，这些有向弧段的集合就构成一类真实的实体或现象。计算机存储了这些有向弧段，因此也就存储了现实世界的地理实体和图形。

当对存储的目标或数据库进行 GIS 操作或分析时，同一层数据可以按类型提取，重要地理实体如县以上境界、县以上居民地、主要河流、主要交通线等可以按实体提取。同一图幅的不同图层的数据可以相互叠合，相邻图幅的各层数据可以拼接和裁切等等。

§3 空间数据结构的类型

在介绍地理空间的表达时已经指出，对现实世界的数据表达可以采用矢量数据模型和栅格数据模型。那么，一旦数据模型确定，必须选择和该模型对应的数据结构来组织实体的数据，最后是选择适合于记录该数据结构的文件格式。

数据结构一般分为基于矢量模型的数据结构和基于栅格模型的数据结构。按照传统的观念，矢量和栅格似乎是两类完全不同性质的数据结构。矢量数据是面向地物的结构，即对于每一个具体的目标都直接赋有位置和属性信息以及目标之间的拓扑关系说明。但是矢量数据仅有一些离散点的坐标，在空间表达方面它没有直接建立位置与地物的关系，如多边形的中间区域是"洞"或"岛"，其间的任何一点并没有与某个地物发生联系。与此相反，栅格数据是面向位置的结构，平面空间上的任何一点都直接联系到某一个或某一类地物。但对于某一个具体的目标又没有直接聚集所有信息，只能通过遍历栅格矩阵逐一寻找，它也不能完整地建立地物之间的拓扑关系。因而，从概念上形成了基于矢量和基于栅格两种类型的系统，分别用于不同的目的。目前，为了设计一种系统能用于多种目的，正在研制一种一体化的数据结构，该数据结构具有矢量和栅格两种结构的特性，称为矢量栅格一体化的数据结构。以下分别介绍上述三种不同类型的数据结构。

3.1 矢量数据结构

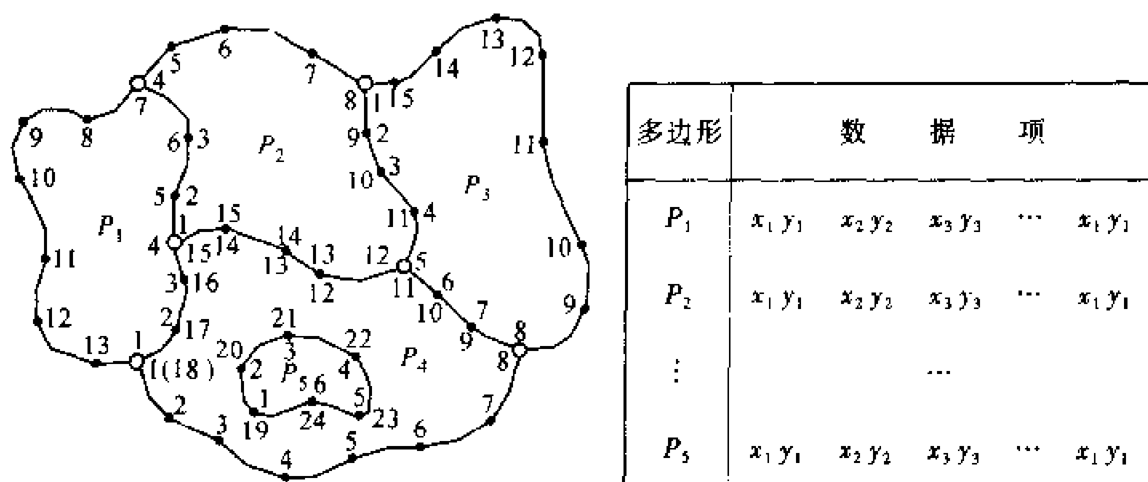
基于矢量模型的数据结构简称为矢量数据结构。矢量也叫向量，数学上称"具有大小和方向的量"为向量。在计算机图形中，相邻两结点间的弧段长度表示大小，弧段两端点的顺序表示方向，因此弧段也是一个直观的矢量。

矢量数据结构是利用欧几里得(Euclid)几何学中的点、线、面及其组合体来表示地理实体空间分布的一种数据组织方式。这种数据组织方式能最好地逼近地理实体的空间分布特征，数据精度高，数据存储的冗余度低，便于进行地理实体的网络分析，但对于多层空间数据的叠合分析比较困难。矢量数据结构分为以下几种主要类型：

1. 简单数据结构

在简单数据结构中，空间数据按照以基本的空间对象(点、线或多边形)为单元进行单独组织(图2-10)，不含有拓扑关系数据，最典型的是面条(Spaghetti)结构。

这种数据结构的主要特点是：



多边形	数　据　项				
P_1	x_1y_1	x_2y_2	x_3y_3	…	x_1y_1
P_2	x_1y_1	x_2y_2	x_3y_3	…	x_1y_1
⋮			…		
P_5	x_1y_1	x_2y_2	x_3y_3	…	x_1y_1

图2-10　Spaghetti数据结构及编码文件

(1) 数据按点、线或多边形为单元进行组织，数据编排直观，数字化操作简单。

(2) 每个多边形都以闭合线段存储，多边形的公共边界被数字化两次和存储两次，造成数据冗余和不一致。

(3) 点、线和多边形有各自的坐标数据，但没有拓扑数据，互相之间不关联。

(4) 岛只作为一个单个图形，没有与外界多边形的联系。

2. 拓扑数据结构

拓扑数据结构包括DIME(对偶独立地图编码法)、POLYVRT(多边形转换器)、TIGER(地理编码和参照系统的拓扑集成)等。它们共同的特点是：点是相互独立的，点连成线，线构成面。每条线始于起始结点(FN)，止于终止结点(TN)，并与左右多边形(LP和RP)相邻接。构成多边形的线又称为链段或弧段，两条以上的弧段相交的点称为结点，由一条弧段组成的多边形称为岛，多边形图中不含岛的多边形称为简单多边形，表示单连通区域；含岛区的多边形称为复合多边形，表示复连通区域。在复连通区域中，包括有外边界和内边界，岛区多边形看作是复连通区域的内边界，复连通区域的内边界多边形对应的区域含有平面上的无穷远点。该数据结构的基本元素如图2-11所示。

在这种数据结构中，弧段或链段是数据组织的基本对象。弧段文件由弧段记录组成，每个弧段记录包括弧段标识码、FN、TN、LP和RP。结点文件由结点记录组成，包括每个结点的结点号、结点坐标及与该结点连接的弧段标识码等。多边形文件由多边形记录组成，包括多边形标识码、组成该多边形的弧段标识码以及相关属性等。

现以图2-11为例，列出拓扑数据结构的弧段文件格式(表2-5)：

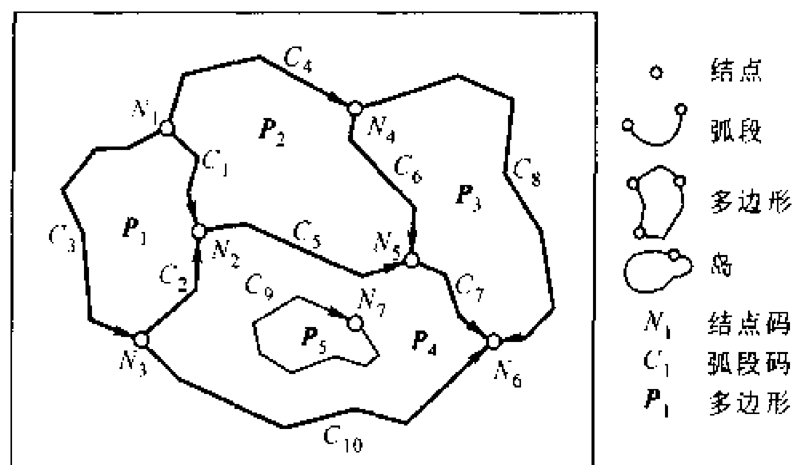


图 2-11 矢量结构图形基本元素

表 2-5 拓扑数据结构的弧段文件构成

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_3	N_2	P_1	P_4
C_3	N_1	N_3	P_1	Φ
C_4	N_1	N_4	Φ	P_2
C_5	N_2	N_5	P_2	P_4
C_6	N_4	N_5	P_3	P_2
C_7	N_5	N_6	P_3	P_4
C_8	N_4	N_6	Φ	P_3
C_9	N_7	N_7	P_4	P_5
C_{10}	N_3	N_6	P_4	Φ

拓扑数据结构最重要的技术特征和贡献是具有拓扑编辑功能。这种拓扑编辑功能，不但保证数字化原始数据的自动查错编辑，而且可以自动形成封闭的多边形边界，为由各个单独存储的弧段组成所需要的各类多边形及建立空间数据库奠定基础。

拓扑编辑功能包括多边形连接编辑和结点连接编辑，前者指顺序连接组成封闭多边形一组线段的编辑，后者指顺序连接环绕某个结点所有多边形的编辑。具体的编辑算法如下：

(1) 多边形连接编辑。例如，设需要对多边形 P_1 进行编辑，其算法过程为：

① 从表 2-5 所示的弧段文件中，检出与当前编辑的多边形 P_1 相关的所有记录：

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_3	N_2	P_1	P_4
C_3	N_1	N_3	P_1	Φ

② 在检出的记录中，计算机检查当前编辑的多边形 P_1 所处的位置：

如果 P_1 位在左多边形位置，将之与位于右多边形位置的多边形号相交换，同时也将该记录的结点号位置作相应的交换；反之，如果当前编辑的多边形 P_1 位于右多边形位置，则该记录的所有数据项顺序不作改变。

按照上述规则，检出的记录变为以下形式：

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_2	N_3	P_4	P_1
C_3	N_3	N_1	Φ	P_1

③ 从经过代码位置转换的记录中，任取一个起结点作为起点，顺序连接各个结点，必要时可对记录的前后顺序作调整，使得连接的结点能自行封闭，如图 2-12 所示。

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_2	N_3	P_4	P_1
C_3	N_3	N_1	Φ	P_1

图 2-12 结点连接形成封闭多边形

如果依照上述顺序连接的结点不能自行闭合，或者出现记录缺损或记录多余等情况，则表示弧段文件有错，必须改正出错的记录。直到所有多边形都经过编辑和改正，再转入结点连接编辑。

(2) 结点连接编辑。例如，设需要对结点 N_2 进行编辑，其算法过程为：

① 从表 2-5 所示的弧段文件中，检出与当前编辑的结点 N_2 相关的所有记录：

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_3	N_2	P_1	P_4
C_5	N_2	N_5	P_2	P_4

② 在检出的记录中，计算机检查当前编辑的结点 N_2 所处的位置：

如果 N_2 位在起结点位置，将之与位于终结点位置的结点号相交换，同时也将该记录的多边形号位置作相应的交换；反之，如果当前编辑的结点 N_2 位于终结点位置，则该记录的所有数据项顺序不作改变。

按照上述规则，检出的记录变为以下形式：

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_3	N_2	P_1	P_4
C_5	N_5	N_2	P_4	P_2

③ 从经过代码位置转换的记录中，任取一个左多边形作为起点，顺序连接各个多边形，同样，必要时可对记录的前后顺序作调整，使得连接的多边形能首尾呼应，如图 2－13 所示。

弧段号	起结点	终结点	左多边形	右多边形
C_1	N_1	N_2	P_2	P_1
C_2	N_3	N_2	P_1	P_4
C_5	N_5	N_2	P_4	P_2

图 2－13　环绕结点 N_2 的多边形连接

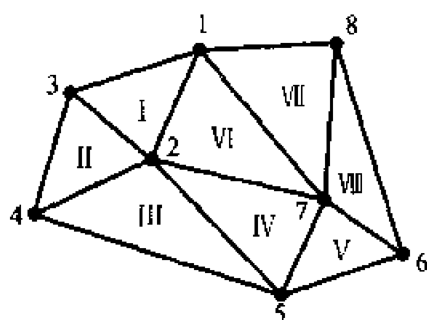
如果依照上述顺序连接的多边形不能首尾呼应，或者出现记录缺损或记录多余等情况，同样也表示弧段文件有错，必须改正出错的记录。直到所有结点都经过编辑和改正，才能将该弧段文件用于结点文件和多边形文件的自动生成以及数据库的建立。

这种拓扑数据结构及其自动编辑功能，已经被许多商品化的 GIS 软件所采用，例如美国的 ARC/INFO GIS 软件等。

3．曲面数据结构

曲面是指连续分布现象的覆盖表面，具有这种覆盖表面的要素有地形、降水量、温度、磁场等。表示和存储这些要素的基本要求是必须便于连续现象在任一点的内插计算，因此经常采用不规则三角网来拟合连续分布现象的覆盖表面，称为 TIN(Triangulated Irregular Network)数据结构(图 2－14)。

这种基于 TIN 的曲面数据结构，通常用于数字地形的表示，或者按照曲面要素的实测点分布，将它们连成三角网，三角网中的每个三角形要求尽量接近等边形状，并保证由最邻近的点构成的三角形，即三角形的边长之和最小。在所有可能的三角网中，狄洛尼(Delaunay)三角网在地形拟合方面表现最为出色，因此常被用于 TIN 的生成。



三角形识别码	相邻三角形			三角形顶点			顶点坐标和特征值		
	Δ1	Δ2	Δ3	1st	2nd	3rd	$X_1\ Y_1\ Z_1$	$X_2\ Y_2\ Z_2$	$X_3\ Y_3\ Z_3$
Ⅰ	Ⅱ	Ⅵ	0	1	3	2			
Ⅱ	Ⅰ	Ⅲ	0	4	3	2			
Ⅲ	Ⅱ	Ⅳ	0	5	4	2			
Ⅳ	Ⅲ	Ⅴ	Ⅵ	7	2	5			
Ⅴ	Ⅳ	Ⅷ	0	6	5	7			
Ⅵ	Ⅰ	Ⅳ	Ⅶ	7	1	2			
Ⅶ	Ⅵ	Ⅷ	0	8	1	7			
Ⅷ	Ⅴ	Ⅶ	0	8	6	7			

图 2-14 不规则三角网及其数据组织

狄洛尼三角形由三个最邻近的点连接而成，这三个相邻点对应的 Voronoi 多边形有一个公共的顶点，此顶点同时也是狄洛尼三角形外接圆的圆心。图 2-15 描述了欧几里得平面上 6 个点的狄洛尼三角网以及 Voronoi 图的对偶，从中可以看出狄洛尼三角网遵守平面图形的欧拉定理：

多面体顶点数 - 边数 + 面数 = 2

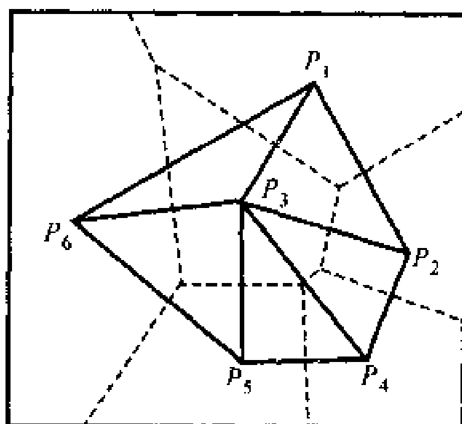


图 2-15 6个平面点集合的狄洛尼三角网

狄洛尼三角网中的每个三角形可视为一个平面，平面的几何特性完全由三个顶点的空间坐标值$(x_i, y_i, z_i)(i=1,2,3)$所决定。每个三角形分别构成一个记录，每个记录的数据项包括：三角形标识码、该三角形的相邻三角形号、该三角形的顶点号以及三个顶点的空间坐标值。Z 轴一般用于表示现象的属性，例如地形高程等。在三角形的任一边上可以进行连续分布现象的线性内插。这种数据结构的相邻三角形信息可以自动生成，而且利用这种相邻三角形信息，便于连续分布现象的顺序追踪和查询检索，例如对地形结构线的追踪，是非常便捷的。利用这种数据结构，可以方便地进行地形分析，如坡度和坡向信息提取，填挖方计算，阴影和地形通视分析，等高线自动生成和 2.5 维显示等。因此，TIN 数据结构被广泛应用于各种地理信息系统，如 ARC/INFO、MGE 等。

3.2 栅格数据结构

基于栅格模型的数据结构简称为栅格数据结构，指将空间分割成有规则的网格，在各个网格上给出相应的属性值来表示地理实体的一种数据组织形式。

在栅格数据结构中，点由一个单元网格表示，其数值与近邻网格值明显不同。线段由一串有序的相互连接的单元网格表示，各个网格的值比较一致，但与邻域的值差异较大。多边形由聚集在一起的相互连接的单元网格组成，区域内部的网格值相同或差异较小，但与邻域网格的值差异较大。

栅格数据结构表示的是二维表面上地理要素的离散化数值，每个网格对应一种属性，其空间位置用行和列标识。网格通常是正方形，有时也采用矩形、等边三角形和六边形。网格边长决定了栅格数据的精度，但是，当用栅格数据来表示地理实体，不论网格边长多细，与原实体特征相比较，信息都有丢失，这是由于复杂的实体采用统一的格网所造成的。一般地，可以通过保证最小多边形的精度标准来确定网格尺寸，使形成的栅格数据既有效地逼近地理实体，又能最大限度降低数据冗余度。

如图 2－16 所示，设研究区域最小图斑的面积为 A，当网格边长为 H 时，该图斑可能丢失；当边长为 $H/2$ 时，该图斑得到很好的表示。所以，合理的网格尺寸为：

$$H = \frac{1}{2}(\min\{A_i\})^{1/2}$$

式中：$i = 1,2,\cdots,n$（区域多边形数）。

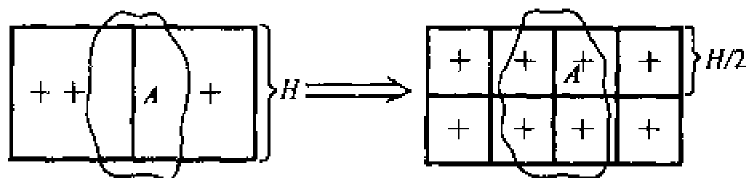


图 2－16 栅格尺寸（H）的确定

按照这个标准建立的栅格数据，其逼近图形的效果（图 2－17），与原来的图形（图 2－18）比较，具有很好的相似性或精度。

栅格数据结构与矢量数据结构相比较，用栅格数据结构表达地理要素比较直观，容易实现多元数据的叠合操作，便于与遥感图像及扫描输入数据相匹配建库和使用等。但是，栅格数据结构的缺点也是很显著的，例如数据精度取决于网格的边长，当网格边长缩小时，网格单元的数量将呈几何级数递增，造成存储空间的迅速增加；由于相邻网格单元属性值的相关性，造成栅格数据的冗余度，特别当表示不规则多边形时，数据冗余度更大；栅格数据对于网络分析和建立网络

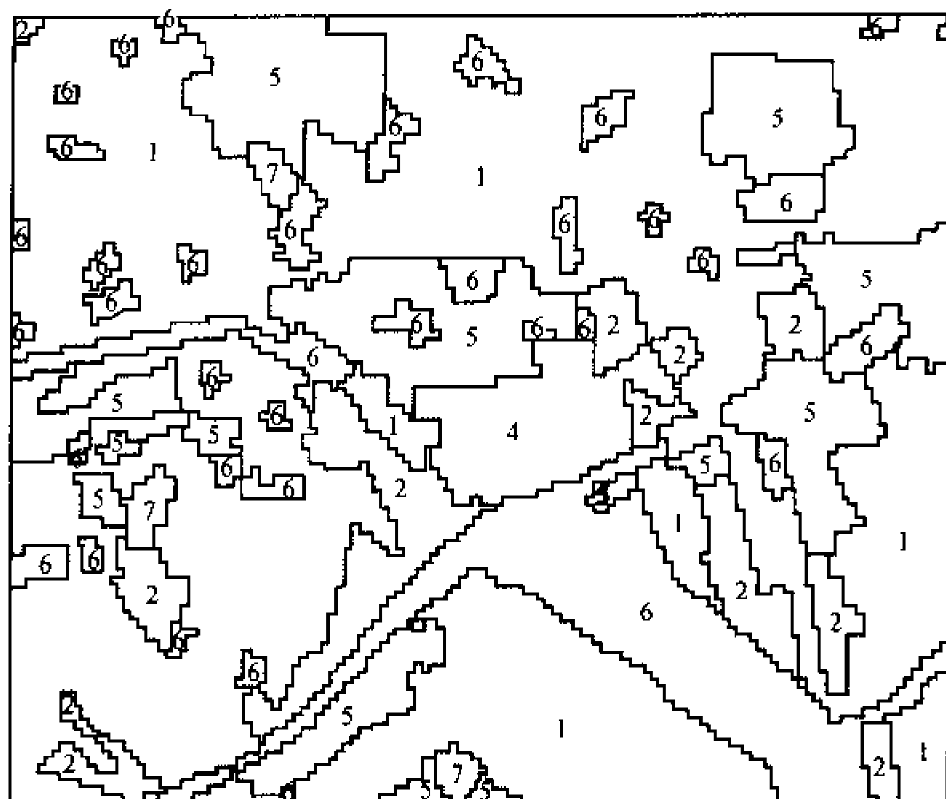


图 2-17　栅格数据逼近的土地利用图

1. 水田　2. 旱地　3. 居民地　4. 湖塘　5. 晒场　6. 湖塘　7. 晒场

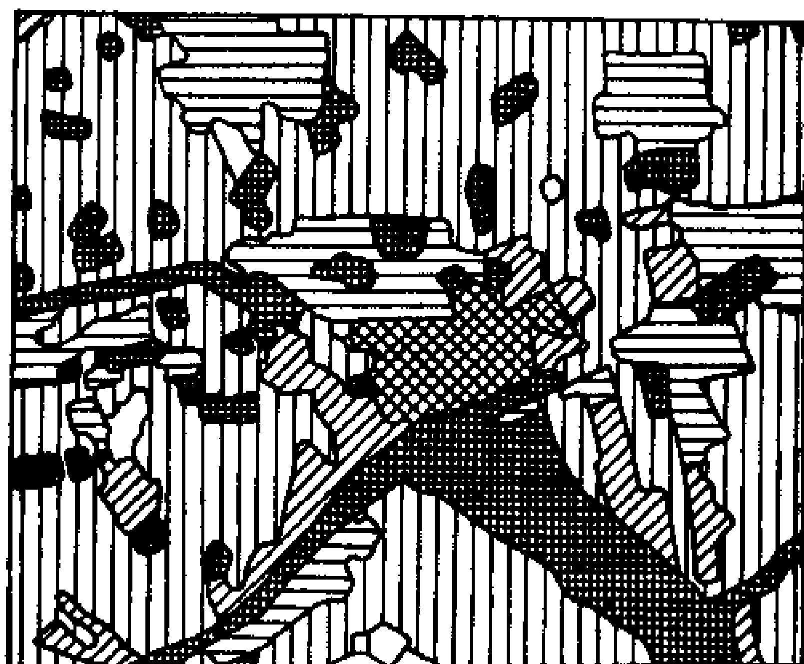


图 2-18　原始的土地利用图

连接关系比较困难等。

栅格数据结构可以分为以下几种主要类型：

1. 栅格矩阵结构

栅格矩阵结构是指一种全栅格阵列的空间数据组织形式(图 2－19)。

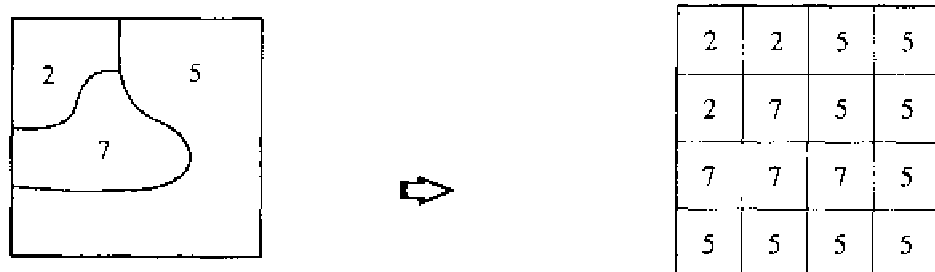


图 2－19 面域栅格矩阵结构

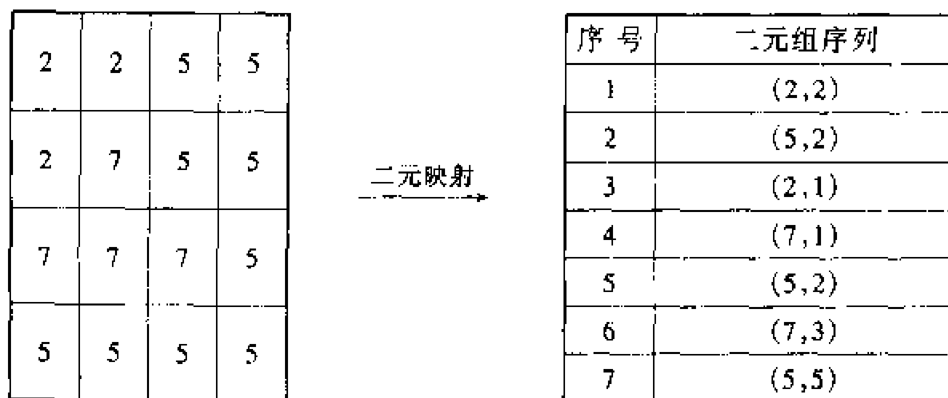
上述数据代表一个 4×4 阶的矩阵，如果矩阵的每个元素用一个双字节表示，则一个图层的全栅格数据所需要的存储空间为 m(行)×n(列)×2(字节)。以一个面积为 100 km^2 的区域为例，如果网格边长取为 1 m，每个网格用一个双字节表示，则一个图层的要素就要占用 200 兆字节的存储空间，对一张图形或一幅图像来说，这是一个相当大的存储容量，而且随着空间分辨率的提高，存贮空间成几何级数递增。因此，栅格数据的压缩是栅格数据结构要解决的重要任务之一。

2. 游程编码结构

游程指相邻同值网格的数量，游程编码结构是逐行将相邻同值的网格合并，并记录合并后网格的值及合并网格的长度，其目的是压缩栅格数据量，消除数据间的冗余。

游程编码结构的建立方法是：将栅格矩阵的数据序列 $X_1X_2\cdots X_n$，映射为相应的二元组序列(A_i，P_i)，$i=1,K$，且 $K\leqslant n$。其中，A 为属性值，P 为游程，K 为游程序号。

例如，将图 2－19 的栅格矩阵结构转换为游程编码结构，如图 2－20 所示。



序号	二元组序列
1	(2,2)
2	(5,2)
3	(2,1)
4	(7,1)
5	(5,2)
6	(7,3)
7	(5,5)

图 2－20 游程编码表示栅格矩阵数据

这种数据结构特别适用于二值图像数据的表示，如图 2-21 所示。

1	1	1	1	1
1	1	0	0	0
0	0	1	1	1
1	1	0	0	0
0	1	1	1	1

二元映射 →

序号	二元组序列
1	(1,7)
2	(0,5)
3	(1,5)
4	(0,4)
5	(1,4)

图 2-21　游程编码表示二值图像数据

游程编码能否压缩数据量，主要决定于栅格数据的性质，通常可通过事先测试，估算图层的数据冗余度 R_r：

$$R_r = 1 - \frac{Q}{m \cdot n}$$

式中：Q 为图层内相邻属性值变化次数的累加和；

m 为图层网格的行数；

n 为图层网格的列数。

当 R_r 的值大于 1/5 的情况下，表明栅格数据的压缩可取得明显的效果。其压缩效果，可由压缩比 $S = n/K$ 来表征，即压缩比的值愈大，表示压缩效果愈显著。

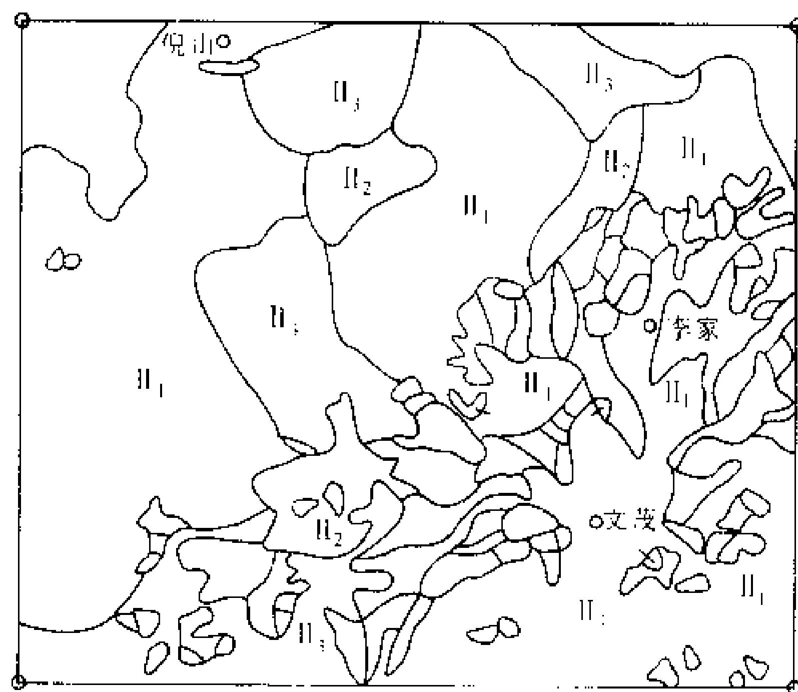


图 2-22　土壤面域分布图

以图2－22为例，当取网格边长为1 mm时，建立的栅格矩阵数据个数为100×120个，每个数据可用一个8位码表示。采用游程长度压缩编码法表示时，建立的游程总数为572个，每个游程用一个双字节整数表示，则压缩比为10∶1，因此大大压缩了直接编码的存储容量，而且信息不丢失。这是一种信息熵保持的地理编码方法，通过解码可以恢复为原始的栅格矩阵格式。

当栅格数据为规则的数字地形高程（即DEM）时，由于这种类型数据的相邻数据具有高度的相关性，可先通过差分映射进行预处理，然后再采用游程长度压缩编码法。其实施过程如图2－23所示。

5	10	10	10	11	12	13	5	5	5
10	20	20	10	10	10	10	10	10	10
20	20	20	60	50	20	20	40	40	40
30	30	20	20	20	20	20	40	40	40
20	20	40	20	20	20	20	40	40	40
20	20	20	20	20	20	20	40	40	40
10	20	50	20	20	20	60	67	40	40
20	10	50	20	20	107	60	20	20	10
10	127	40	40	40	20	10	10	10	10
40	40	40	20	20	20	20	20	20	100

⇒

5	5	0	0	1	1	1	－8	0	0
10	10	0	－10	0	0	0	0	0	0
20	0	0	40	－10	－30	0	20	0	0
30	30	－10	0	0	0	0	20	0	0
20	0	20	－20	0	0	0	20	0	0
20	0	0	0	0	0	0	20	0	0
10	10	30	－30	0	0	40	7	－27	0
20	－10	40	－30	0	87	－47	－40	0	－10
10	117	－87	0	0	－20	－10	0	0	0
40	0	0	0	－2	0	0	0	0	－80

⇓

(5,2)　(0,4)　(1,7)　(－8,8)　(0,10)

(10,2)　(0,3)　(－10,4)　(0,10)

(20,1)　(0,3)　(40,4)　(－10,5)　(－30,6)　(0,7)　(20,8)　(0,10)

(30,1)　(0,2)　(－10,3)　(0,7)　(20,8)　(0,10)

(20,1)　(0,2)　(20,3)　(－20,4)　(0,7)　(20,8)　(0,10)

(20,1)　(0,7)　(20,8)　(0,10)

⋮　⋮　⋮　⋮

(10,1)　(117,2)　(－87,3)　(0,5)　(－20,6)　(－10,7)　(0,10)

(40,1)　(0,4)　(－20,5)　(0,9)　(－80,10)

图2－23　游程长度压缩编码法

栅格数据经过压缩处理，得到游程编码数据序列，为了提高系统对这些数据的访问效率，通常采用索引顺序文件的方法来组织数据（图2－24）。当由位置参数访问其属性特征时，利用逻辑顺序和逻辑地址的关系，很快在索引文件中找到指向数据文件欲访栅格的指针，并求出其逻辑地址，就能找到该栅格的属性。

反之，由属性查访其分布位置时，通过对数据文件进行遍历扫描，由于这种扫描是在大大少于原始栅格数据的编码数据上进行的，因此，其遍历的速度也是很快的。

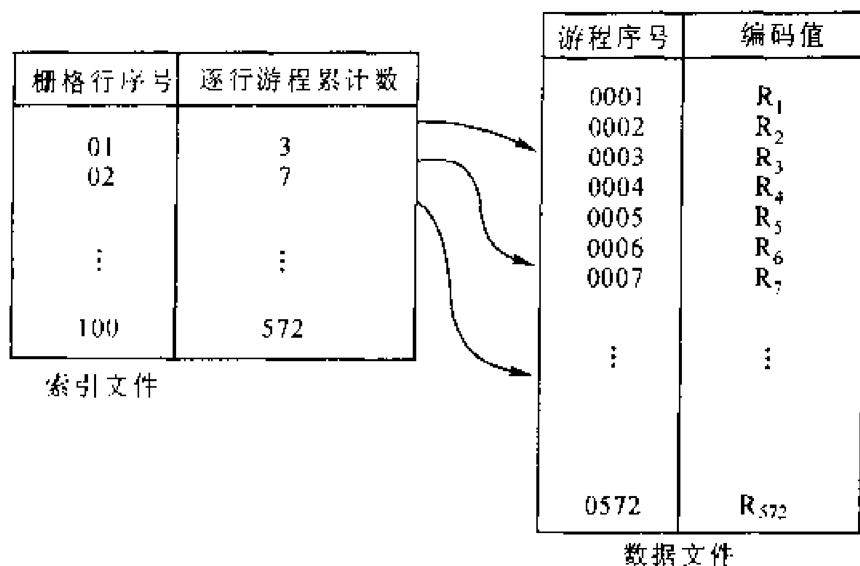


图 2-24 游程编码数据的组织

3. 四叉树数据结构

空间要素的区域信息可用四叉树数据结构存储，这种数据结构的原理可以表述为：将空间区域按照四个象限进行递归分割（$2^n \times 2^n$，且 $n \geqslant 1$），直到子象限的数值单调为止。凡数值（特征码或类型值）呈单调的单元，不论单元大小，均作为最后的存储单元。这样，对同一种空间要素，其区域格网的大小，随该要素分布特征而不同。如图 2-25 所示，图 2-25(a)为区域（晕线部分）四分划的过程；图 2-25(b)为该区域对应的四叉树，其中树根代表整个区域，树的每个结点有四个儿子或者空，为空的结点称为叶结点，叶结点对应于区域分割时数值单调的子象限。

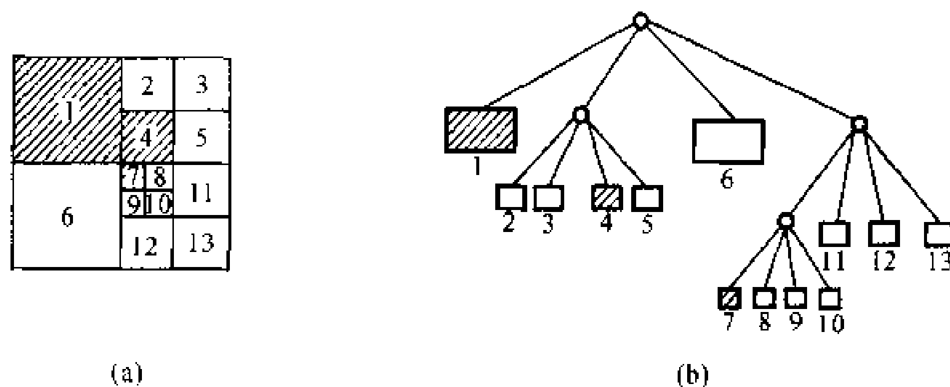


图 2-25 四叉树压缩编码表示法

建立四叉树有两种方法，即自上而下方式（“top-down”）和自下而上方式（“bottom-up”）。自上而下方式是先检测全区域，其值不单调时再四分划，直到数值或内容单调为止。例如，对一个由 $n \times n(n=2\times k, k\geqslant 1)$ 的栅格方阵组成的区域 P（图 2-26），它的四个子象限（P_a，P_b，P_c，P_d）分别为：

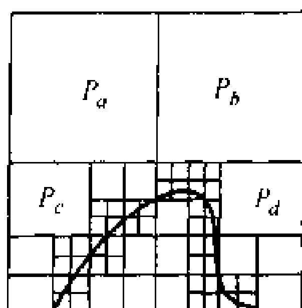


图 2-26 区域 P 子象限的表示

$$P_a=\left\{P[i,j]:1\leqslant i\leqslant \frac{1}{2}n, 1\leqslant j\leqslant \frac{1}{2}n\right\};$$

$$P_b=\left\{P[i,j]:1\leqslant i\leqslant \frac{1}{2}n, \frac{n}{2}+1\leqslant j\leqslant n\right\};$$

$$P_c=\left\{P[i,j]:\frac{n}{2}+1\leqslant i\leqslant n, 1\leqslant j\leqslant \frac{1}{2}n\right\};$$

$$P_d=\left\{P[i,j]:\frac{n}{2}+1\leqslant i\leqslant n, \frac{n}{2}+1\leqslant j\leqslant n\right\};$$

再下一层的子象限分别为：

$$P_{aa}=\left\{P[i,j]:1\leqslant i\leqslant \frac{1}{4}n, 1\leqslant j\leqslant \frac{1}{4}n\right\};$$

$$\vdots$$

$$P_{ba}=\left\{P[i,j]:1\leqslant i\leqslant \frac{1}{4}n, \frac{n}{2}+1\leqslant j\leqslant \frac{3}{4}n\right\};$$

$$\vdots$$

$$P_{dd}=\left\{P[i,j]:\frac{3}{4}n+1\leqslant i\leqslant n, \frac{3}{4}n+1\leqslant j\leqslant n\right\};$$

其中，a、b、c、d 标号分别表示西北（WN）、东北（EN）、西南（WS）、东南（ES）四个子象限。根据这些表达式可以求得任一层的某个子象限在全区的行列位置，并对这个位置范围内的网格值进行检测。若数值单调，就不再细分。按照这种方法，可以完成整个区域四叉树的建立，但是需要大量的运算，因为一些已经划分好的区块会被重复检测，特别是当 $n \times n$ 的矩阵比较大，区域内容要素比较复杂时，用这种方法建立四叉树的速度比较慢。自下而上方式是按图 2-27 的顺序扫描。

1	2	5	6	17	18	21	22
3	4	7	8	19	20	23	24
9	10	13	14	25	26	29	30
11	12	15	16	27	28	31	32
33	· · ·						
35	· · ·						

图 2-27 自下而上建立四叉树的顺序

设栅格数据为 $A(n,n)(n=2\times k,k\geqslant 1)$，则先检测 $A(1,1)$、$A(1,2)$、$A(2,1)$、$A(2,2)$，然后是 $A(1,3)$、$A(1,4)$、$A(2,3)$、$A(2,4)$等等。若4个格网值相同，则合并；反之，作为4个叶结点记录。依此逐层向上，直到最后生成根结点。这就是网格数据的四叉树化。

四叉树每个结点通常存储6个量，4个子结点指针、一个父结点指针（根结点的父指针为空，叶结点的子指针为空）和一个结点值，这就是通常说的常规四叉树（CQT）。另一种存储方法是每个结点只存储3个量：地址、深度和结点值，即称为线性编码四叉树（LQT）。以下详细讨论这两种四叉树建立的方法。

(1) 线性四叉树编码的存储算法。线性四叉树编码是计算好每个网格单元的地址，其计算公式为 ADDRES$(I、J)=2*I_B+J_B$，式中 I_B、J_B 分别为行(I)和列(J)的二进制数形式。若区域是 $2^K\times 2^K$ 的矩阵，这样生成的地址由 K 个数字组成。如 64×64 的区域，其地址是一个有6位数字的数。一般在计算机文件中要用四个字节才能存储它。当这样算出的地址按照由小到大排序后，正是“bottom-up”（自下而上）法扫描检测的顺序。程序依次判断相邻四个网格值是否完全相同。若不完全相同，则作为四个叶结点记录下来；若完全相同，则将它们合而为一，并将其中一个网格值赋给它，它的地址为原四个单元中第一个单元的地址，为节省存储量，中间结点不予存储。

由于栅格数据常常不恰好是 $2^n\times 2^n$ 的方阵，为了能对不同行列数的栅格数据进行四叉树编码，在设计程序时，对不是 $2^n\times 2^n$ 的部分以0补足。（这样确定 n 的值，使得 2^n 大于或等于行、列数中的大者，取合乎要求的最小的自然数为 n）。在建树时，对于补足部分生成的叶结点可不存储，这样存储量并不会增加。考虑到许多栅格数据是用游程编码存储的，可直接应用游程数据建立四叉树。

这种线性四叉树只存储三个值，比常规四叉树节省存储量。由于记录结点地址，既能直接找到其在四叉树中的走向路径，也可以换算出它在整个栅格区域内的行列位置。

(2) 常规四叉树编码的存储算法。常规四叉树也采用自下而上的方法建立，对栅格按一定的顺序进行检测。这种方法除了要记录叶结点外，还要记录中间结点。每记录四个结点时，即生成其父结点。

如果在检测时不进行合并，建立所谓的丰满四叉树，第 k 级有4个方块，所以丰满四叉树结点共有：

$$\text{NUM}=\sum_{k=0}^{n}4k=\frac{4^{n+1}-1}{3}\approx\frac{4}{3}4^n\text{ 个}$$

即比栅格矩阵的单元数目多33.3%，但是通过合并，其实际结点数远远不

会这么多。

常规四叉树虽然增加了一些存储量，却换取了处理上的简便和灵活。而且，如果一个图像的栅格矩阵很大，存储和处理整幅图很困难时，采用常规四叉树编码存储法，可以根据需要将之分成4块、16块、64块等等，每块分别进行四叉树编码存储。需要时，可将它们合并成一棵树，合并的方法对常规四叉树来说十分容易。例如，如果要将相邻4块的四叉树合并成一棵树，只要将4块的四叉树当做新树中的4棵子树，4棵子树的根指向一个共同的父亲，即重新生成一个共同的根结点就行了。

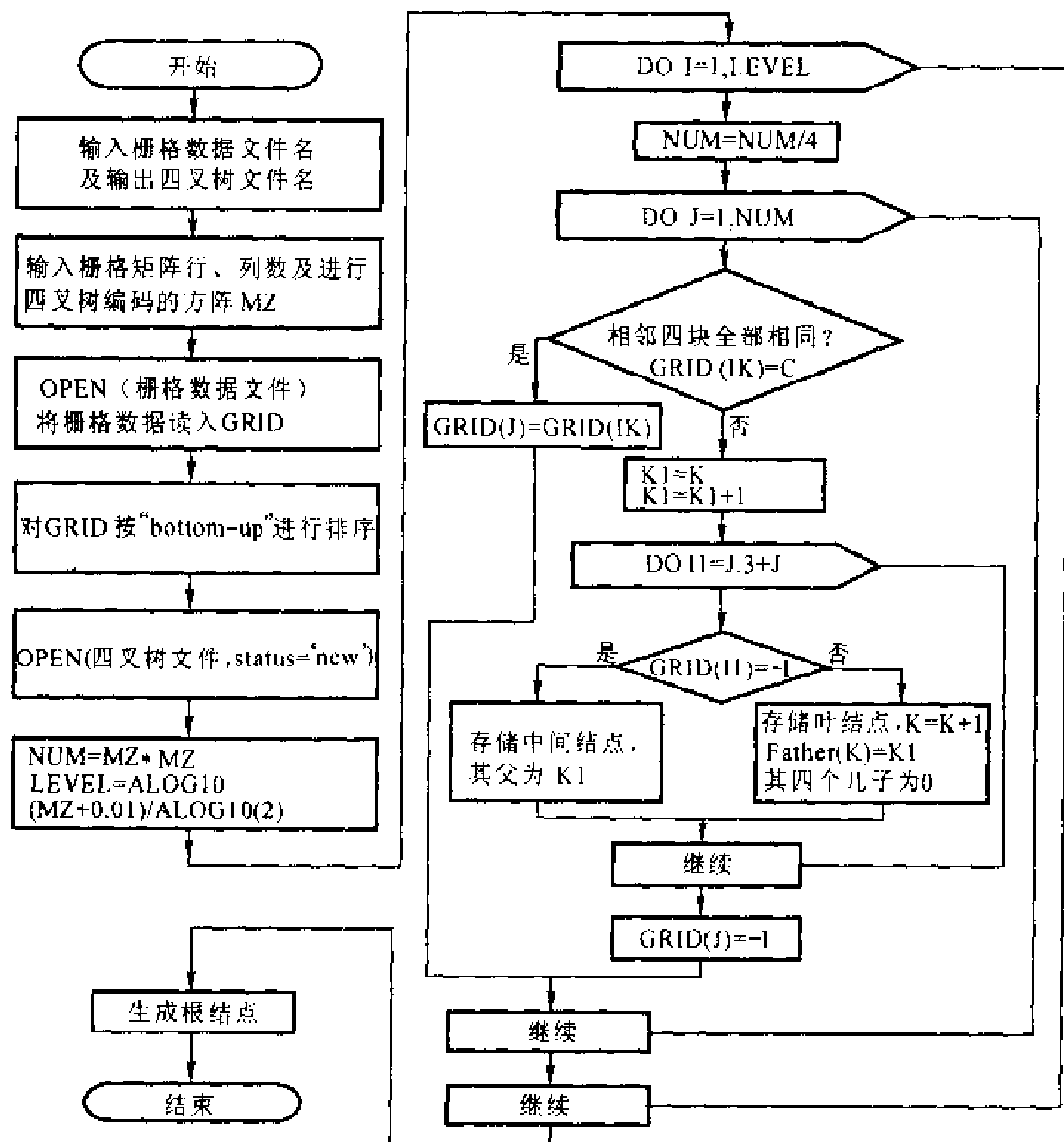


图2－28 常规四叉树编码存储程序框图

常规四叉树对于数据检索，多要素叠合分析，求物体间的空间关系等操作都

十分方便。

下面给出常规四叉树编码存储的程序框图(图2-28),其中数组GRID存储栅格数据。

数据Father存储结点的父指针。

4. 八叉树和十六叉树结构[①]

(1) 三维和四维数据结构的提出。前面介绍的数据结构都是二维的,然而在有些信息系统中,需要有真三维的空间数据结构。例如矿山开采中的地下资源埋藏和采矿巷道的空间分布,如果用二维的坐标体系就根本无法很好表达。此外,矿山空间目标往往随时间不断变化着,这就提出了空间和时间信息系统的问题。

在时间信息系统中不考虑时间,是把时间看作不变的常数,即为当前的时间。而在时间和空间信息系统中,则把时间看作有过去、现在和将来的可变值。这种系统中,空间和时间是不可分割的信息,并起着同样重要的作用。

我们首先用三维来定义空间目标,在同一坐标系统下,用四维数据来定义时间和空间数据。根据这种方案,任何目标都可以由其坐标对($\{s\}, t$)惟一确定。这里$\{s\} = \{x, y, z\}$定义空间数据,而t定义时间数据。对每一个三维坐标数据(x, y, z),必定有而且仅有一个时间t值与之相对应,但反之则不然。

为了表示三维数据和四维数据,较好的数据结构方式是在四叉树基础上发展起来的八叉树和十六叉树结构。

(2) 八叉树结构及其编码。八叉树结构是从四叉树结构直接发展而来的,其原理就是将空间区域不断地分解为八个同样大小的子区域(即将一个六面的立方体再分解为八个相同大小的小立方体),分解的次数越多,子区域就越小,一直到同一区域的属性单一为止。按从下而上合并的方式来说,就是将研究区空间先按一定的分辨率将三维空间划分为三维栅格网,然后按规定的顺序每次比较8个相邻的栅格单元,如果其属性值相同则合并,否则就记盘。依次递归运算,直到每个子区域均为单值为止。

八叉树同样可分为常规八叉树和线性八叉树。常规八叉树的结点要记录十个值,即八个指向子结点的指针,一个指向父结点的指针和一个属性值(或标识号)。而线性八叉树则只需要记录叶结点的地址码和属性值。因此,它的主要优点是,一是节省存储空间,因为只需对叶结点编码,节省了大量中间结点的存储。每个结点的指针也免除了,而从根到某一特定结点的方向和路径的信息隐含在定位码之中,定位码数字的个位数显示分辨率的高低或分解程度;其次,线性八

① 据郭达志等. 地理信息系统基础与应用. 北京:煤炭工业出版社,1998

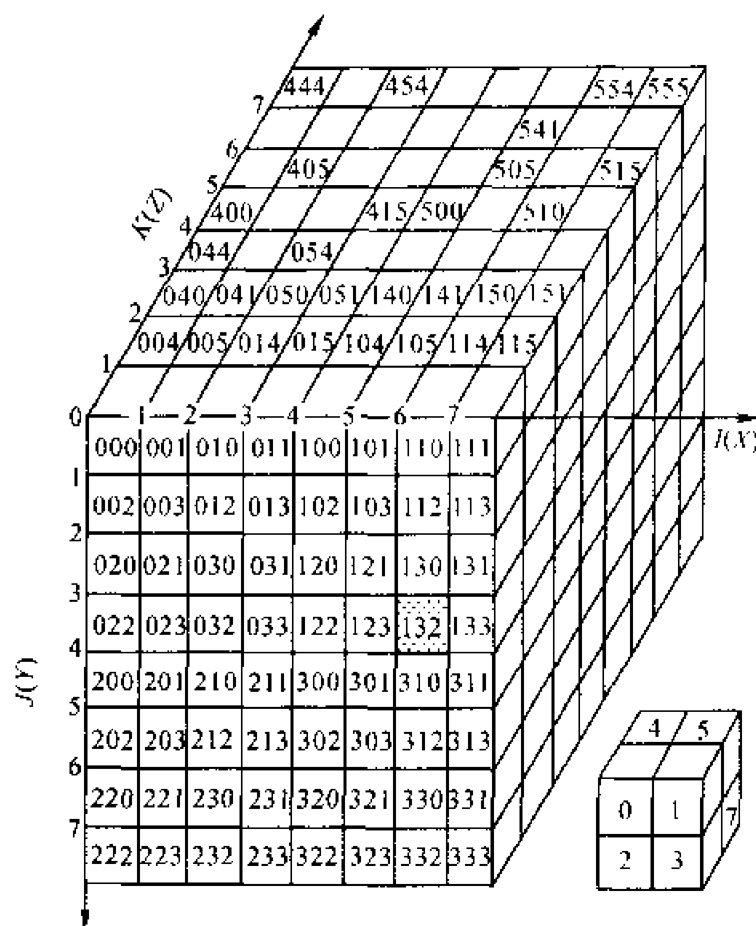


图 2-29 八叉树编码图解（$n=3$）

叉树可直接寻址，通过其坐标值则能计算出任何输入结点的定位码（称编码），而不必实际建立八叉树，并且定位码本身就是坐标的另一种形式，不必有意去存储坐标值。若需要的话还能从定位码中获取其坐标值（称解码）；第三，在操作方面，所产生的定位码容易存储和执行，容易实现集合、相加等组合操作；此外，如果应用任务很大致使在核心存储器中不能容纳所有定位码时，也可以将定位码安排在 B 树中，以便划分成许多页面，并存储在外围设备中。

图 2-29 表示一种线性八叉树的生成。通过对八叉树及其编码方案的分析研究，可总结出下列一些自动计算定位码的有用规则：

① 给定分辨率 n，也即确定了坐标系统的大小。每个坐标轴的取值范围从 0 到 2^n-1。在图 2-29 中的分辨率等于 3，则三个坐标轴的坐标范围为从 0 到 7。

② 每个八叉树点的编码采用的形式如图 2-30 所示。每个位置上的 q_i 是（0,1,2,…,7）八个数之一，q_i 的个数取决于分辨率 n。在本例中定位码以三个八进制数表示。

③ 对整个编码按 Z 字形的方式进行，其方向取决于如何选择第一、二、三和

第四维坐标。

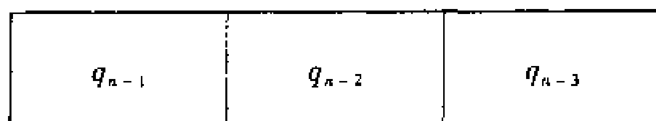


图 2-30 编码方式

基于第一个规则，坐标(X,Y,Z)已知的八叉树点的编码可用下式表示：

$$\left.\begin{aligned} X &= c_{n-1}2^{n-1} + c_{n-2}2^{n-2} + \cdots + c_0 \\ Y &= d_{n-1}2^{n-1} + d_{n-2}2^{n-2} + \cdots + d_0 \\ Z &= e_{n-1}2^{n-1} + e_{n-2}2^{n-2} + \cdots + e_0 \end{aligned}\right\} \qquad (2-1)$$

式中：系数 c_i、d_i 和 $e_i$$(i=0,\cdots,n-2,n-1)$的取值为 0 或 1，可通过比较方程式两边的值来确定。

根据第二、三规则的观点，q_i 值可由下式确定：

$$q_i = e_i 2^2 + d_i 2^1 + c_i (i = n-1, n-2, \cdots, 0) \qquad (2-2)$$

取 $n=3$，$(X,Y,Z)=(6,3,1)$为例，则八叉树编码的表达式为

$$6 = 1.2^{3-1} + 1.2^{3-2} + 0$$

$$3 = 0.2^{3-1} + 1.2^{3-2} + 1$$

$$1 = 0.2^{3-1} + 0.2^{3-2} + 1$$

因此

$$c_2 = 1, c_1 = 1, c_0 = 0$$

$$d_2 = 0, d_1 = 1, d_0 = 1$$

$$e_2 = 0, e_1 = 0, e_0 = 1$$

q_i 的值为

$$q_2 = 0.2^2 + 0.2^1 + 1 = 1$$

$$q_1 = 0.2^2 + 1.2^1 + 1 = 3$$

$$q_0 = 1.2^2 + 1.2^1 + 0 = 6$$

因此，按给定的八叉树节点所产生的定位码为 136。

线性八叉树的地址码有八进制和十进制数两种。十进制地址码亦称 Morton 码。由于 Morton 码是自然数码，所以可以将二维数组转化成以 Morton 码为下标的一维数组。线性八叉树的自然数编码的表达式为

$$\begin{aligned} N = {} & 2^0(c_0 8^0 + c_1 8^1 + \cdots + c_{n-1} 8^{n-1}) \\ & + 2^1(d_0 8^0 + d_1 8^1 + \cdots + d_{n-1} 8^{n-1}) \\ & + 2^2(e_0 8^0 + e_1 8^1 + \cdots + e_{n-1} 8^{n-1}) \end{aligned} \qquad (2-3)$$

式中：$c_0, d_1, c_{n-1}, d_0, d_1, d_{n-1}$和 e_0, e_1, e_{n-1}分别是行号和列号二进制化后由低位到高位的权。

当图形、图像恢复时，可由 N 码经逆变换为下列十进制的行、列号：

$$I_1 = \sum_{k=n-1}^{k=0} \mathrm{MOD}(T_K, 2) \cdot 2^k \tag{2-4}$$

其中 $T_K = \mathrm{INT}(N_K/8^k)$

当 $k < n-1$ 时，$N_K = \mathrm{MOD}(N_{K+1}, T_{K+1} \cdot 8^{k+1})$；

当 $k = n-1$ 时，$N_K = N$。

$$I_2 = \sum_{k=n-1}^{k=0} \mathrm{MOD}(T_K, 2) \cdot 2^k \tag{2-5}$$

其中 $T_K = \mathrm{INT}(N_K/8^k \cdot 2)$

当 $k < n-1$ 时，$N_K = \mathrm{MOD}(N_{K+1}, T_{K+1} \cdot 8^{k+1} \cdot 2)$；

当 $k = n-1$ 时，$N_K = N$。

$$I_3 = \sum_{k=n-1}^{k=0} \mathrm{MOD}(T_K, 2) \cdot 2^k \tag{2-6}$$

其中 $T_K = \mathrm{INT}(N_K/8^k \cdot 2^2)$

当 $k < n-1$ 时，$N_K = \mathrm{MOD}(N_{K+1}, T_{K+1} \cdot 8^{k+1} \cdot 2^2)$；

当 $k = n-1$ 时，$N_K = N$。

(3) 线性 2^m 叉树的自然数编码及其收敛性分析。基于线性四叉树和线性八叉树的数据结构，可以推想，对于一个四维的时空目标也可以用线性十六叉树来表示。因此推求线性 2^m 叉树的自然数编码的通式，不仅有理论意义，还便于指导实际应用。

根据以上的推导，可以得到线性 2^m 叉树自然数编码的一般公式为

$$N = \sum_{i=1}^{m} 2^{i-1} \sum_{k=0}^{n-1} \mathrm{MOD}(I_{ik}, 2) \cdot (2^m)^k \tag{2-7}$$

其中

$$I_{ik} = \mathrm{INT}(I_{i(k-1)}/2) \qquad (k > 0)$$

式中：m 为研究目标的维数；

I_{ik}是 m 维欧氏空间中第 i 个坐标轴的十进制坐标。

我们知道图像分辨率 n 决定了栅格坐标大小，在式(2-7)中给出了二进制化的最大取余迭代次数就是图像分辨率大小。而在同一图像分辨率下，m 维欧氏空间中不同的坐标取余迭代收敛次数为

$$n' = \mathrm{INT}(\log_2 I_{i0}) + 1 \tag{2-8}$$

表明取余迭代速度是由坐标 I_{i0}决定的，因此线性 2^m 叉树自然数编码的实用公

式为

$$N = \sum_{i=1}^{m} 2^{i-1} \sum_{k=0}^{INT(\log_2 I_{i0})} MOD(I_{ik},2) \cdot 2^{mk} \qquad (2-9)$$

其中　$$I_{ik} = INT(I_{i(k-1)}/2 \qquad (k>0)$$

m 维图像恢复时,进行逆变换:

$$I_i = \sum_{k=n-1}^{k=0} MOD(T_k,2) \cdot 2^k (i = 1,\cdots,m) \qquad (2-10)$$

其中　$T_k = INT(N_k/(2^m)^k \cdot 2^{i-1}) = INT(N_k/2^{mk+i-1}$

当 $k < n-1$ 时,$N_K = MOD(N_{K+1}, T_{k+1} \cdot (2^m)^{k+1} \cdot 2^{i-1}) = MOD(N_{K+1}, T_{K+1}, 2^{mk+m+i-1})$;

当 $k = n-1$ 时,$N_K = N_0$。

同样,为了提高逆变换式(2-10)取余迭代收敛速度,应选确定 k 的取余迭代初始值,即满足条件

$$\frac{N}{2^{mk+i-1}} \qquad (2-11)$$

有　$$k = INT\left[\frac{1}{m}(\log N + 1 - i)\right] \qquad (2-12)$$

因此实用时,式(2-10)中 k 的初值用式(2-12)取代。

利用式(2-10)和式(2-12)进行图像恢复的逆变换时,取余迭代是否在 $k=0$ 处收敛呢?很明显要满足

$$MOD(T_k,2) \equiv 0 \qquad (2-13)$$

恒成立,于是有 $mk+i-1 \leqslant -1$。即

$$k \leqslant -\frac{i}{m} \qquad (1 \leqslant i \leqslant m) \qquad (2-14)$$

只有当 $i=m$ 时,$k \leqslant -1$。因此在取余迭代式(2-10)中,I_i 值一定会在 $k=0$ 处收敛。

作为上述公式应用的一个例子,图2-31为由一维时间和三维空间复合形成的四维时空数据结构的线性十六叉树自然数编码图。

按上述公式可将四维时空坐标系中每个超体素的时空位置转换为线性 2^4 叉树自然数编码,如位于(1,3,0,1)处的超体素的编码 $N=43$。这样十进制的自然数编码 N 自然地对应一维数组下标,四维图像超体素合并运算时,依次将数组中 2^4 个相邻元素进行合并操作。四维图像恢复时,利用上述逆变换为4个坐标值。

线性十六叉树构成的数学模型可扩展如下:

(1) 任何十六叉树的坐标可用下式表示:

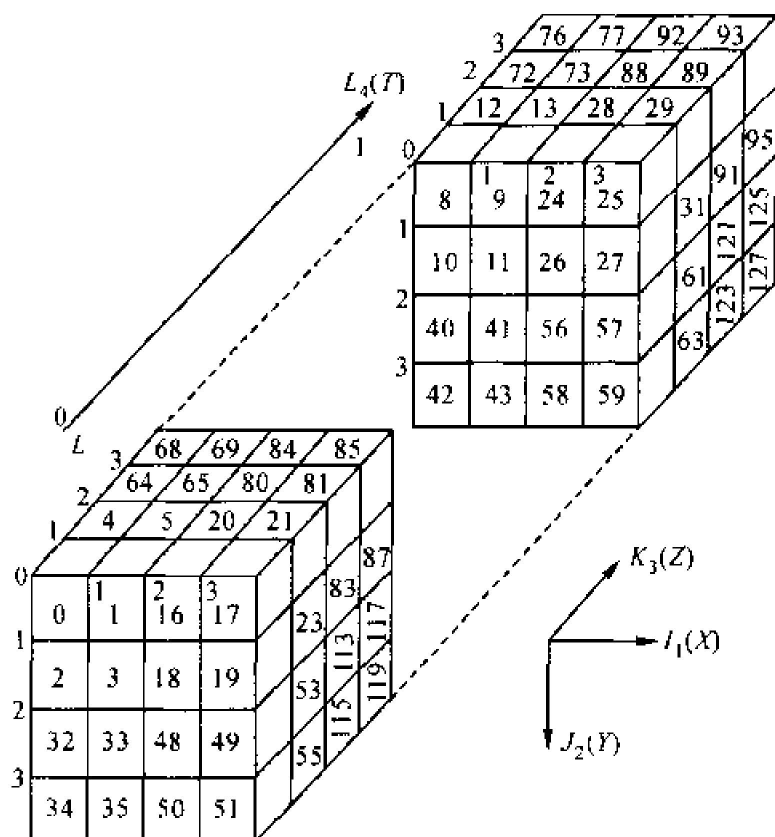


图 2-31 线性十六叉树自然数编码图解($n=2$)

$$\left.\begin{aligned} X &= c_{n-1}2^{p-1} + c_{n-2}2^{p-2} + \cdots + c_0 \\ Y &= d_{n-1}2^{p-1} + d_{n-2}2^{p-2} + \cdots + d_0 \\ Z &= e_{n-1}2^{p-1} + e_{n-2}2^{p-2} + \cdots + e_0 \\ T &= f_{n-1}2^{p-1} + f_{n-2}2^{p-2} + \cdots + f_0 \end{aligned}\right\} \qquad (2-15)$$

(2) 定位码的十六位数字 q_i 为

$$q_i = f_i 2^2 + e_i 2^2 + d_i 2^1 + c_i$$

$(i = n-1, n-1, \cdots, 0)$

举例来说,空间和时间坐标为$(X,Y,Z,T)=(3,1,2,1)$,则利用上面公式得到地址码为 $5B$,即如图 2-31 中后面那个立方体中灰色立方单元。而前面那个灰色单元的坐标为(2,1,0,0),它的地址码为 12。

3.3 矢量与栅格一体化数据结构

1. 矢量与栅格一体化的基本概念

新一代集成化的地理信息系统,要求能够统一管理图形数据、属性数据、影像数据和数字高程模型(DEM)数据,称为四库合一。关于图形数据与属性数据的统一管理,近年来已取得突破性的进展,不少 GIS 软件商先后推出各自

的空间数据库引擎(SDE),初步解决了图形数据与属性数据的一体化管理。而矢量与栅格数据,按照传统的观念,认为是两类完全不同性质的数据结构,当利用它们来表达空间目标时,对于线状实体,人们习惯使用矢量数据结构。对于面状实体,在基于矢量的GIS中,主要使用边界表达法,而在基于栅格的GIS中,一般用元子空间填充表达法。由此,人们联想到对用矢量方法表示的线状实体,是不是也可以采用元子空间填充法来表示,即在数字化一个线状实体时,除记录原始取样点外,还记录所通过的栅格。同样,每个面状地物除记录它的多边形边界外,还记录中间包含的栅格。这样,既保持了矢量特性,又具有栅格的性质,就能将矢量与栅格统一起来,这就是矢量与栅格一体化数据结构的基本概念。

为了建立矢量与栅格一体化数据结构,要对点、线、面目标数据结构的存储要求作如下的统一约定:

(1) 对点状目标,因为没有形状和面积,在计算机内部只需要表示该点的一个位置数据及与结点关联的弧段信息。

(2) 对线状目标,它有形状,但没有面积,在计算机内部需用一组元子来填满整个路径,并表示该弧段相关的拓扑信息。

(3) 对面状目标,它既有形状,又有面积,在计算机内部需表示由元子填满路径的一组边界和由边界组成的紧凑空间。

由于栅格数据结构的精度较低,需利用细分格网的方法,来提高点、线和面状目标边界线的数据表达精度。如图2-32中,在有点、线目标通过的基本格网内,再细分成256×256个细格网。当精度要求较低时,也可以细分成16×16个细格网。为使数据格式一致,基本格网和细分格网都采用线性四叉树的编码方法,将采样点和线性目标与基本格网的交点用两个Morton码表示(均用十进制Morton码,简称M码)。其中,M_1表示该点(取样点或附加的交叉点)所在的基本格网地址码,M_2表示该点对应的细分格网的Morton码,即M_1和M_2是将同一对X、Y坐标转换成的两个Morton码。例如,$X = 210.00$ m,$y = 12.23$ m,当基本格网的边长取为10 m,在每个弧段通过的基本格网内再细分为256×256个细格网时,可得出$M_1 = 275$,$M_2 = 2\,690$。

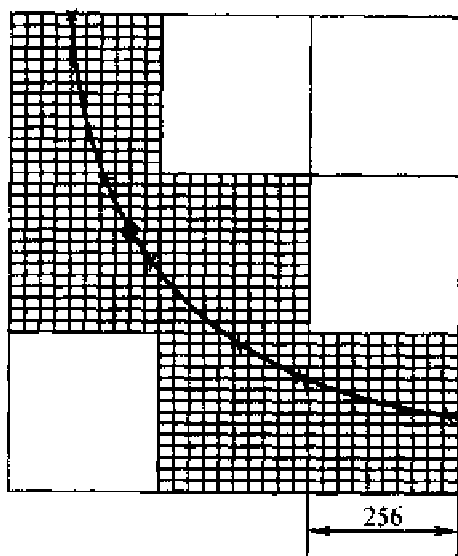
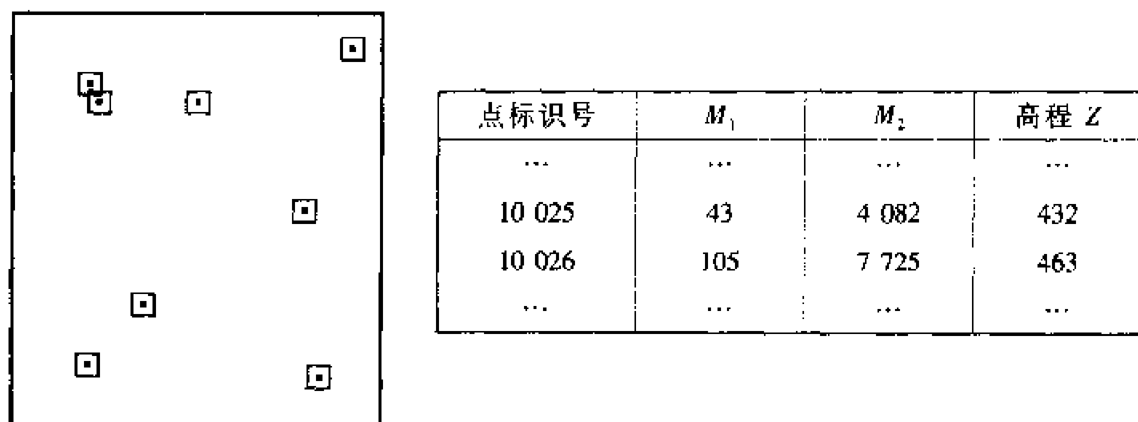


图2-32　细分格网示意图

2. 矢量与栅格一体化数据结构设计

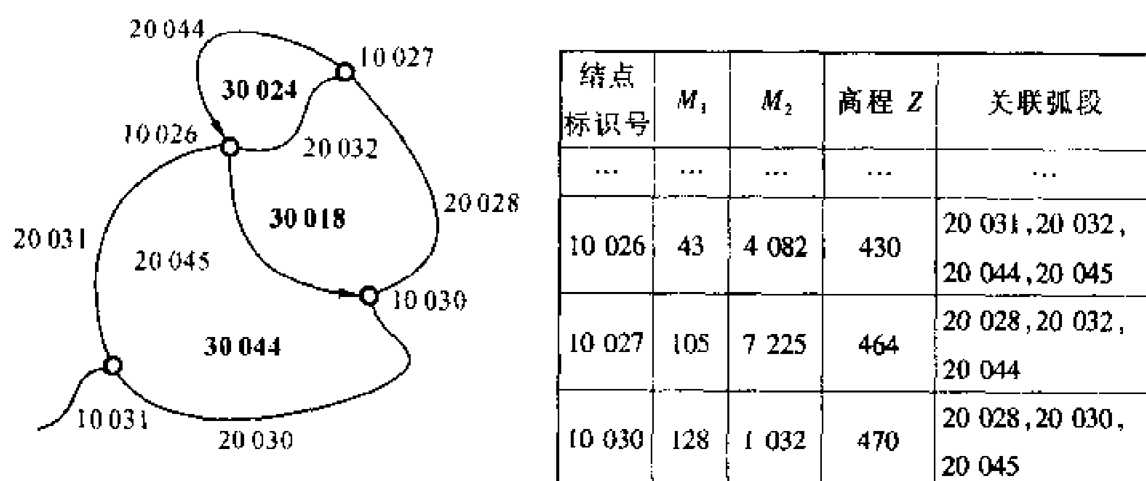
(1) 点状目标和结点的数据结构。根据基本约定,点状目标和结点只有位

置，没有形状和面积，不必将点状地物作为一个覆盖层分解成四叉树，只要将点坐标化为两个地址码 M_1 和 M_2，而不管整个构形是否为四叉树。这种结构简单灵活，便于点的插入和删除等操作。其结构如图 2－33 和图 2－34 所示。



点标识号	M_1	M_2	高程 Z
…	…	…	…
10 025	43	4 082	432
10 026	105	7 725	463
…	…	…	…

图 2－33　点状目标及其数据结构



结点标识号	M_1	M_2	高程 Z	关联弧段
…	…	…	…	…
10 026	43	4 082	430	20 031，20 032，20 044，20 045
10 027	105	7 225	464	20 028，20 032，20 044
10 030	128	1 032	470	20 028，20 030，20 045

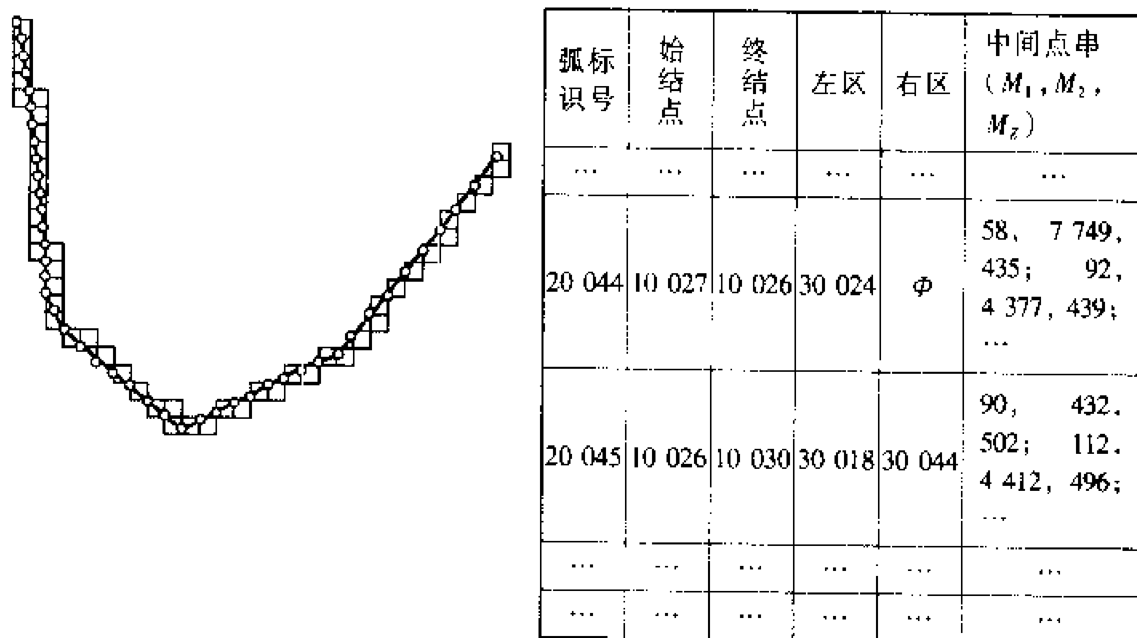
图 2－34　结点及其数据结构

(2) 线状目标和弧段数据结构。根据基本约定，线状目标只要将其通过的栅格地址全部记录下来即可，由于一个线状地物可能由几个弧段组成，所以应先建立弧段的数据文件。其数据结构如图 2－35 所示。

其中，始结点和终结点是该弧段的两个端点（结点），它们与结点数据结构联接，可以建立起弧段与结点之间的拓扑关系。中间点串包含了原始取样点（已转换成用 M_1、M_2 表示）和该弧段路径通过的所有格网边的交点位置码。

(3) 面状目标的数据结构。根据基本约定，面状目标应包含边界和边界所包围的整个区域。边界由弧段组成，关联弧段构成多边形区域，通过关联弧段与弧段数据结构的连接，可以建立多边形与弧段之间的拓扑关系。面状目标中间包含的栅格采用线性四叉树或二维行程数据结构表示。面状目标的数据结构如

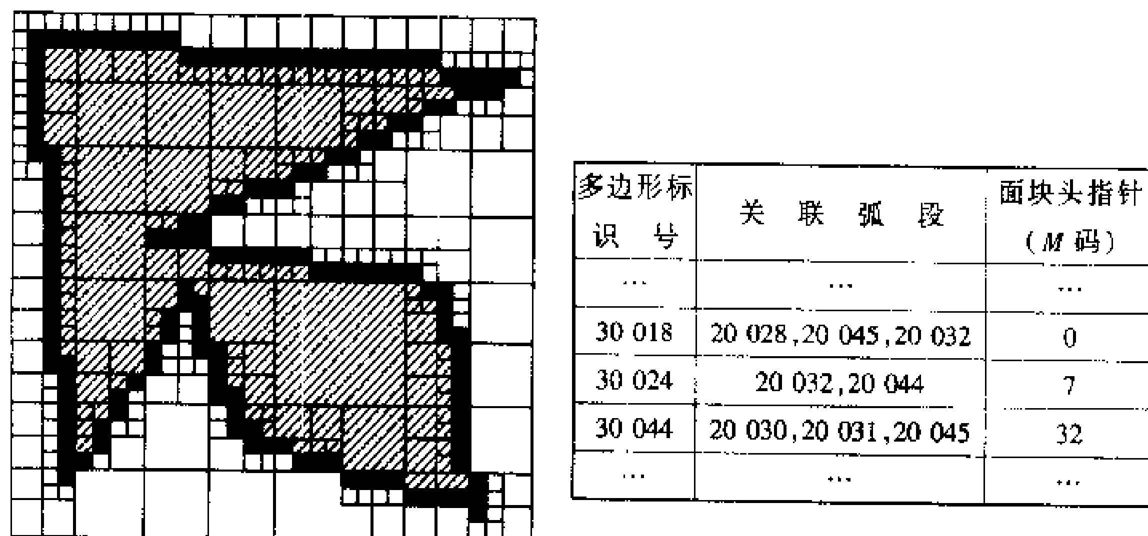
图 2-36 所示。



弧标识号	始结点	终结点	左区	右区	中间点串（M_1，M_2，M_Z）
…	…	…	…	…	…
20 044	10 027	10 026	30 024	Φ	58, 7 749, 435; 92, 4 377, 439; …
20 045	10 026	10 030	30 018	30 044	90, 432, 502; 112, 4 412, 496; …
…	…	…	…	…	…
…	…	…	…	…	…

图 2-35　弧段及其数据结构

其中，面块头指针是指向该目标的下一个子块的记录或地址码，并在最后指向该目标的对应属性值。



多边形标识号	关联弧段	面块头指针（M 码）
…	…	…
30 018	20 028，20 045，20 032	0
30 024	20 032，20 044	7
30 044	20 030，20 031，20 045	32
…	…	…

图 2-36　面状目标及其数据结构

通过带指针的二维行程表（表 2-6），可以用循环指针将属于同一类目标（例如土壤覆盖层的红壤）的叶结点链接起来（图 2-37），形成面向目标或具体地物的面域结构，而且只要进入第一块，就可以顺着指针直接提取该目标的所有子块，避免像栅格矩阵那样，为了查询某一个目标而遍历整个矩阵，从而大大加

快了查询速度。

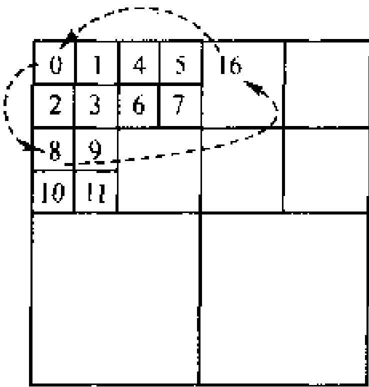


图 2-37 链接面块的指针

表 2-6 带指针的二维行程表

二维行程 M码	循环指针 属性值	二维行程 M码	循环指针 属性值
0	8	32	38
7	12	36	39
8	16	38	40
12	17	39	4
16	0	40	2
17	36	48	8
25	48	⋮	⋮

这里规定一个覆盖层(layer)形成一个二维行程表,若干个覆盖层(例如土地利用、土壤等)则形成若干个二维行程表,以便进行GIS的叠合操作和分析。每个层可用一颗四叉树或一个二维行程编码表表示。叶结点的值可以是属性值或目标的标识号。具体的检索过程是:面状目标标识号→搜索边界弧段→面块头指针→所有面块及其属性值。因此,面状目标数据结构的基础是弧段数据结构及带指针的二维行程编码文件。

3.4 矢量与栅格数据结构的比较

如前所述,矢量数据结构与栅格数据结构是GIS常用的两种数据结构,但它们各有不同的优点和缺点,如表2-7所示。

表 2-7 矢量与栅格数据结构的比较

	优 点	缺 点
矢量数据结构	1. 便于面向现象(土壤类、土地利用单元等)的数据表示 2. 数据结构紧凑、冗余度低 3. 有利于网络分析 4. 图形显示质量好、精度高	1. 数据结构复杂 2. 软件与硬件的技术要求比较高 3. 多边形叠合分析比较困难 4. 显示与绘图成本比较高
栅格数据结构	1. 数据结构简单 2. 空间分析和地理现象的模拟均比较容易 3. 有利于与遥感数据的匹配应用和分析 4. 输出方法快速,成本比较低廉	1. 图形数据量大 2. 投影转换比较困难 3. 栅格地图的图形质量比较低些 4. 现象识别的效果不如矢量方法

§4 空间数据结构的建立

空间数据结构的建立是指根据确定的数据结构类型，形成与该数据结构相适应的GIS空间数据，为空间数据库的建立提供物质基础。

数据结构的建立，通常是融合在系统设计和系统实施过程中，其主要内容和建立过程如图2-38所示。

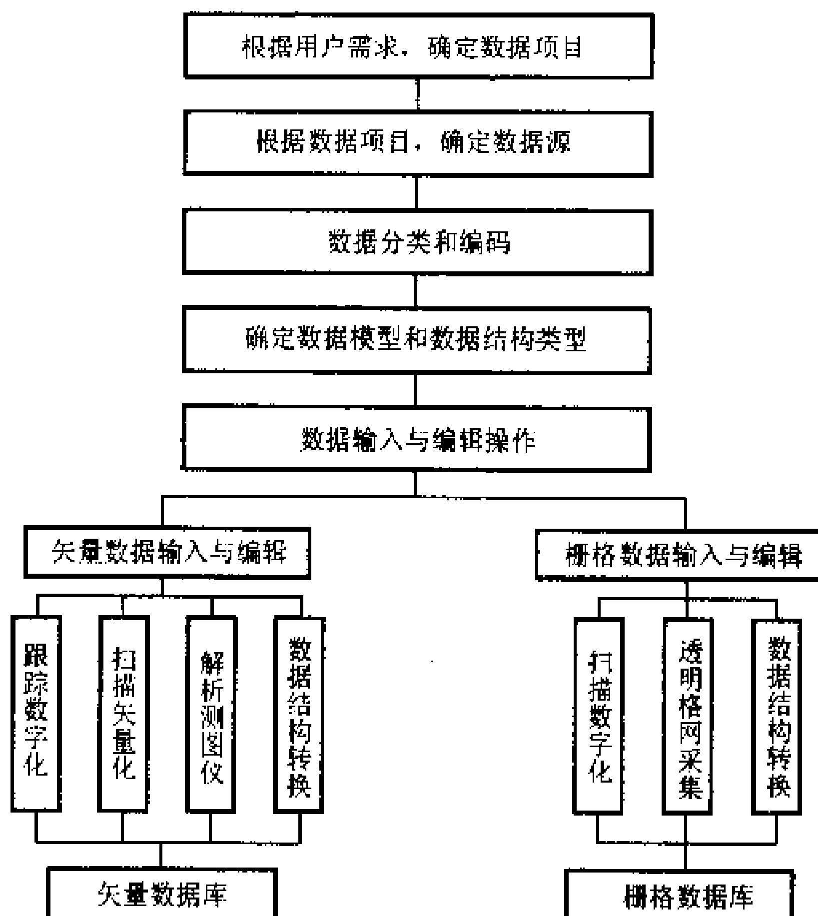


图2-38 数据结构建立的基本过程

4.1 系统功能与数据间的关系

现代地理信息系统数据模式的一个重要特征是数据与功能之间具有密切的联系。例如，以城市信息系统为例，一般需要的原始数据包括十大类：基础底图、环境数据、公用设施、工程平面图、地块图、街区类型数据、区域统计数据、交通统计数据、街道网文件数据和区域境界线数据。其中每一类数据都包括一系列数据实体，每种数据实体由许多数据项组成，例如环境数据包括土壤、地质、植被、地形、水文、坡度和其他地理实体，而每一种地理实体又包含有空间坐标、地理分

类、面积统计等等。确定这些类型的数据是由系统的功能所决定的。例如利用这些专题图的数据是为了研究土地的适宜性和承载力，确定潜在的矿产和森林资源，以及分析可能的自然灾害，包括洪水、地震、塌方、不宜开发的区域等。再如交通统计数据指连接结点之间的道路路线，包括道路通行质量、事故多发地段、人口、住宅条件、垃圾箱数目等，主要用来进行城市交通路线的评价，以及作为城市规划和管理的依据。城市信息系统的各种功能及所需要的数据之间的关系，如表 2－8 所示。

表 2－8　系统功能与数据间的关系（据 jack Dangermond 等，1984）

系统功能＼相关数据	基础数据	环境数据	公用设施数据	工程平面图	地块图数据	街区类型数据	区域统计数据	交通统计数据	街道网文件	区域境界线数据
地产收买和转让		×			×	×	×			×
土地执照发放		×			×	×				×
验收检查	×	×		×	×	×		×		
提供法律咨询	×				×	×		×		
发放许可证										
街道定名										
开展区划	×	×	×	×	×	×		×		
建立街道地址					×			×		
完成事务报告		×			×		×			×
实施调度					×	×				
设置交通路线		×			×		×	×	×	×
交通事故分析		×	×					×	×	
公用设施布设	×	×	×	×	×	×	×	×	×	×
区域划分					×	×	×			
依法进行城市分区		×	×		×	×	×	×	×	×
土地利用规划		×			×	×	×	×	×	×
工程设计	×	×	×	×	×	×	×	×	×	×
地图绘制	×	×	×	×	×	×		×		
名称查询										
完成税收					×	×	×			
邮务管理					×	×	×			
人才资源分配							×			×
公用设施管理			×	×	×	×		×	×	
财产清查管理	×	×	×	×	×	×	×	×	×	×
自然资源管理		×			×	×	×			×
烟草控制										
地图管理	×	×	×		×	×	×			×

续表

系统功能 \ 相关数据	基础数据	环境数据	公用设施数据	工程平面图	地块图数据	街区类型数据	区域统计数据	交通统计数据	街道网文件	区域境界线数据
制图管理			×	×	×			×	×	
数据库管理	×	×			×	×	×	×	×	×
道路开发					×	×				
传播公共信息					×	×	×		×	
回答公众咨询		×			×	×	×	×	×	×

4.2　空间数据的分类和编码

1. 空间数据的分类

空间数据的分类，是指根据系统功能及国家规范和标准，将具有不同属性或特征的要素区别开来的过程，以便从逻辑上将空间数据组织为不同的信息层，为数据采集、存储、管理、查询和共享提供依据。

在进行具体分类时，首先根据图形原则，将空间数据分为点、线、面三种类型；其次是对象原则，例如河流和道路，虽然它们同为线状要素，但是属于不同的地理对象，应当作为不同的数据存储层。

以我国基础地理信息数据分类为例，分为测量控制点、水系、居民地、交通、管线与垣栅、境界、地形与土质和植被八个大类，然后再依次细分为小类、一级类、二级类等。

2. 空间数据的编码

空间数据的编码，是指将数据分类的结果，用一种易于被计算机和人识别的符号系统表示出来的过程。编码的结果是形成代码。代码由数字或字符组成，或由它们共同组成的混合码。

编码的目的，是用来提供空间数据的地理分类和特征描述，同时为了便于地理要素的输入、存储、管理，以及系统之间数据交换和共享的需要。

我国基础地理信息数据，其分类代码由六位数字组成。代码结构如图2-39。

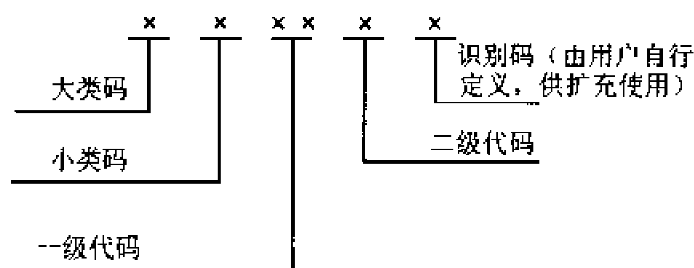


图2-39　国家基础地理信息数据的分类码结构

由于分类系统包含着多级的含义,因此使用的特征码必须是统一拟定的编码系统,并符合各专业学科的分类分级体系。要使拟定的特征码能为多用途数据库提供足够的参考信息,同时又使冗余信息最少;要描述地理要素间的相互关系。例如,美国地质调查局(USGS)制定的《数字线划图形标准》采用七位数字的代码结构,其中前三位为主码,后四位为子码。主码的前二位数字用以惟一定义要素的类别,主码的第三位数字是子码的解释位:如果为零,表示子码是要素的分类码;如果非零,表示子码是要素的参数值或称为参数属性代码,例如地形高程、河流长度等等。子码的第一位数字通常为零,其余三位数字标识要素的图形类型(点状、线状或面状图形)、分级分类(计曲线、间曲线或助曲线等)和其他特征(洼地、冰川、雪原,或河流的左岸、右岸等)。因此,该编码系统具有分类逻辑性强,代码信息量丰富,便于进行要素间关系的推理判别等特点。我国国内行政区,按国家颁布的 GB-2260-80 编码,其中省(直辖市、自治区)占三位,其余三位由用户自行定义,最多为十位。世界或地区拟采用 GB-2659-81 编码。对各专业要素,采用相应专业部门制定的统一编码系统。

4.3 矢量数据的输入与编辑

矢量数据的输入过程实际上是产生和矢量数据结构相适应的 GIS 空间数据的过程,即将经分类和编码的地理要素的图形,转换为一系列 x, y 坐标,然后将这些坐标记录按照确定的数据结构格式,加入到线段或标识点的计算机文件中去。适合产生这种空间数据的图形数字化方法包括:手扶跟踪数字化仪输入、扫描矢量化输入、解析测图仪数据输入及其他数据传输和转换输入等。

例如,当使用手扶跟踪数字化仪输入时,其操作过程为:

(1) 原图准备。即准备数字化原图,并将原图内容转绘到不易引起变形的聚脂薄膜上,检查原图内容的完整性,分别标识原图的图幅角点、内部的控制点等。

(2) 输入初始化参数。包括数字化的阈值(数字化两点间的最短距离),图幅四角点的理论坐标,图幅长度(length)和宽度(width),以及图名、图号、比例尺、地图投影等。

(3) 输入数字化图幅四角点的坐标和经标识的控制点坐标。

(4) 输入数字化图幅内的图形,直到形成一个数字化信息存储层。

(5) 检查和修改数字化错误。通过荧屏或绘图显示,检查结点不匹配、假结点、悬挂结点、线段过长或过短、标识点遗漏等各种错误,并予以改正。

(6) 建立拓扑关系和输入属性。属性数据一般采用键盘输入,作为一个顺序文件,经编辑处理后,再转存到数据库的相应文件或表格中。

(7) 检查和修改拓扑错误。例如线段缺失或多余、多边形不能自行闭合等,

将出错信息打印出来，并逐一予以改正。

(8) 检查和修改属性表错误。例如属性与图形不能一一对应或者出错等，并予以改正。

空间数据编辑的目的是为了消除数字化过程中引入的错误，以及将数字化数据重新组织以便得到便于进一步处理和使用的格式。数字化过程引入的错误可能包括空间点位和线段的丢失或重复、线段过长或过短、区域中心识别码的遗漏，以及结点代码和区域属性码不能符合拓扑一致性的要求等等。重新组织数字化数据的内容取决于输入文件和输出结果的性质，一般包括数字化文件的连接，将图幅数据分割为更小的和更便于处理的各块，将空间数据按照要求或者按照属性进行重新组织，以及将数据组织为可以互换的格式等。

根据 D.F.Marble 教授设计的层状和模块式的数字化数据编辑原理，空间数据编辑的方法和过程如图 2-40 所示。

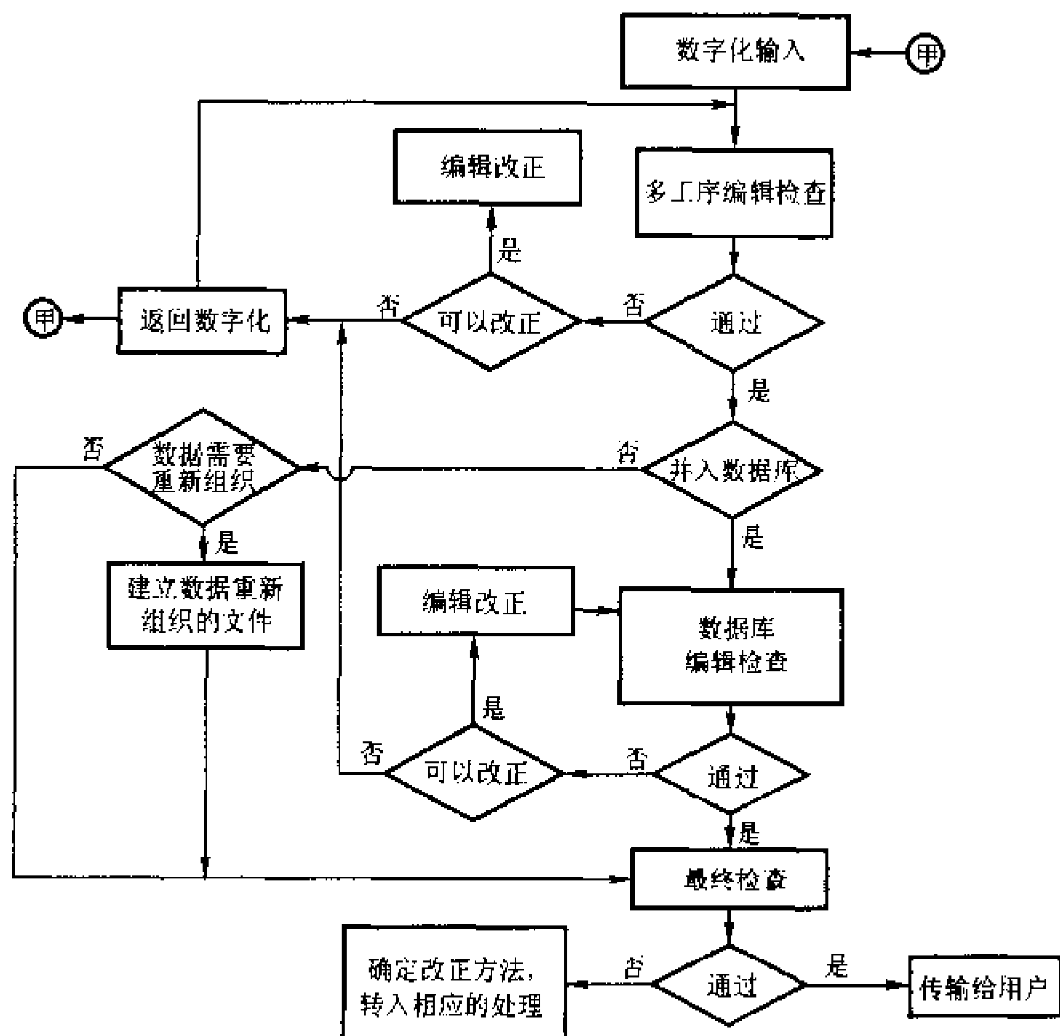


图 2-40 空间数据编辑过程

因此,当原始的数字化文件输入编辑站,就要进行多工序的编辑检查,它包括:目视检查,用以检查一些显著的错误,例如线段的丢失、线段过长等;机器检查,用以进行数字化数据拓扑一致性的逻辑检验,将弧连接成多边形以及进行数字化限差的检查等;图形检查,指对着数字化原图,进行图形的详细检查,通常是利用透光桌,将数字化文件输出的地图覆盖在原始图上,通过人工进行图形的逐一比较。如果精度要求较高,可以将两种图形放大,然后再进行比较。其具体过程如图 2-41 所示。

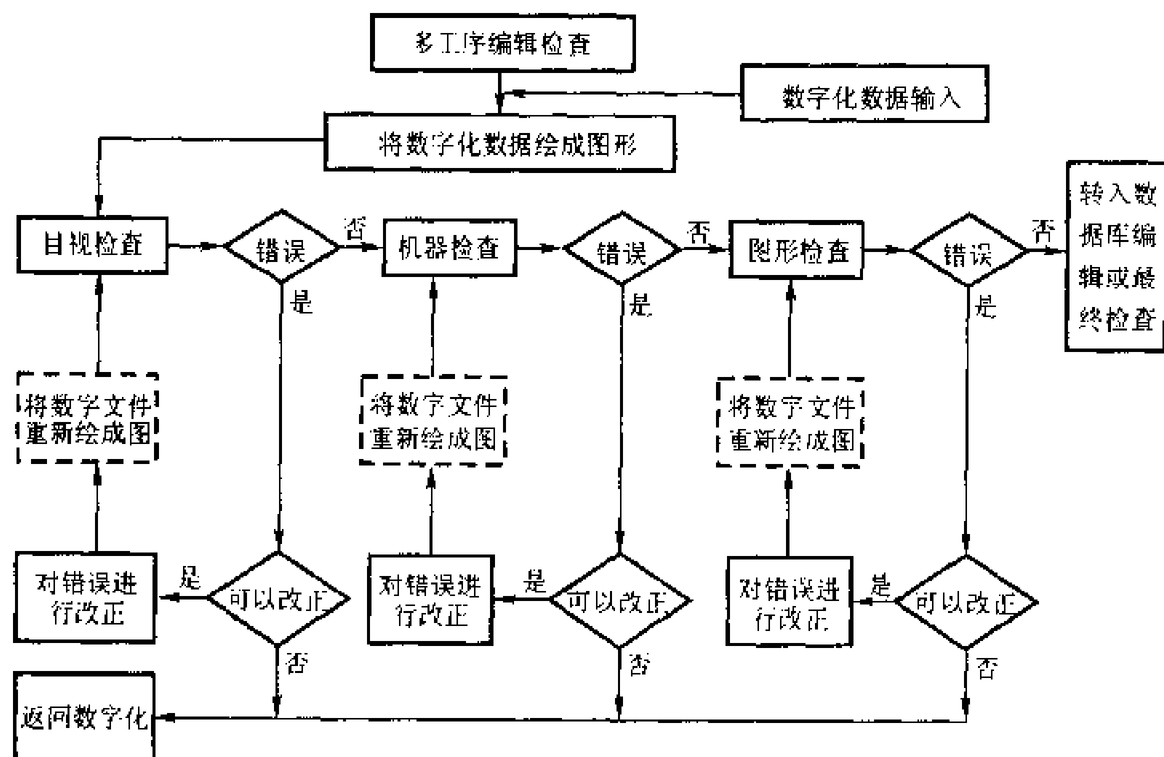


图 2-41 多工序编辑检查的内容

在编辑阶段,当需要将几幅图并为较大的数据库时,需要按照数据库的编辑要求进行空间数据的编辑处理,其中主要是边沿匹配(见第 3 章)。为保证边沿匹配的顺利进行,对并入数据库的数字文件的所有实体要进行认真的检查。进行数据合并时,有时还需要对数字文件进行重新格式化(见第 3 章)。其具体编辑过程见图 2-42。

当文件或数据库的全部编辑任务完成以后,首先要将这种净化后的数据存入新的存储介质,然后再提供应用,同时关闭空间数据编辑站。但这里必须指出,在空间数据编辑阶段只能消除数字化过程中产生的操作错误和图形的一些明显误差,对于因图纸变形而产生的误差和坐标变换等,要通过几何纠正和投影转换来处理,具体方法详见第 3 章的介绍。

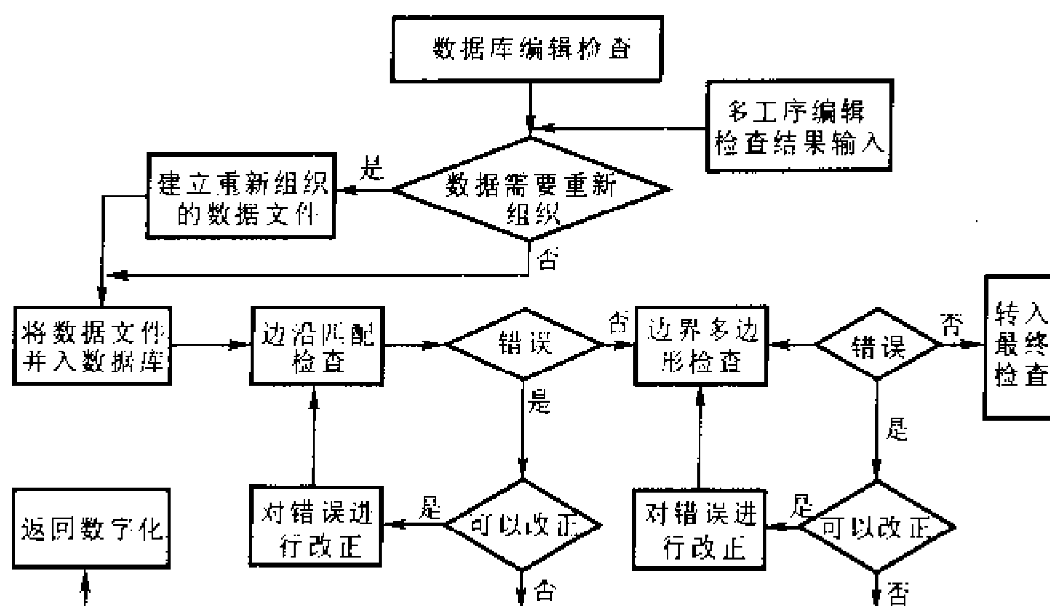


图2-42　数据库编辑检查

4.4　栅格数据的输入与编辑

同样，栅格数据的输入过程是产生和栅格数据结构相适应的GIS空间数据

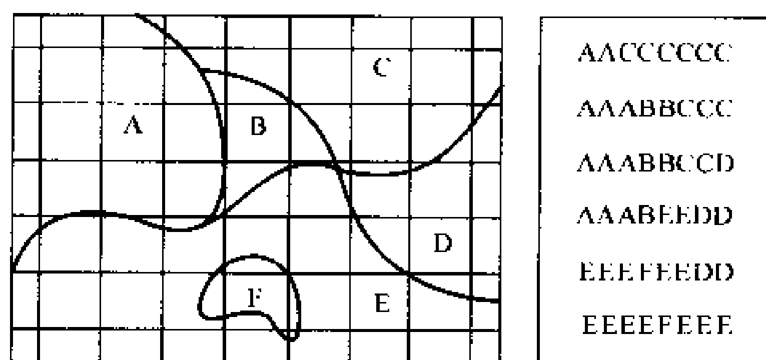


图2-43　单位格网交点归属法

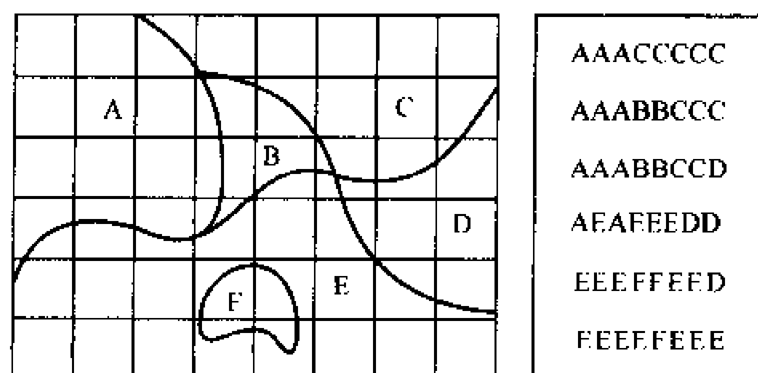


图2-44　单位格网面积占优法

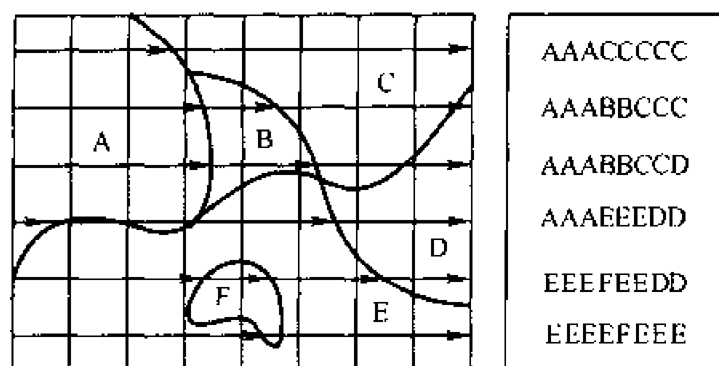
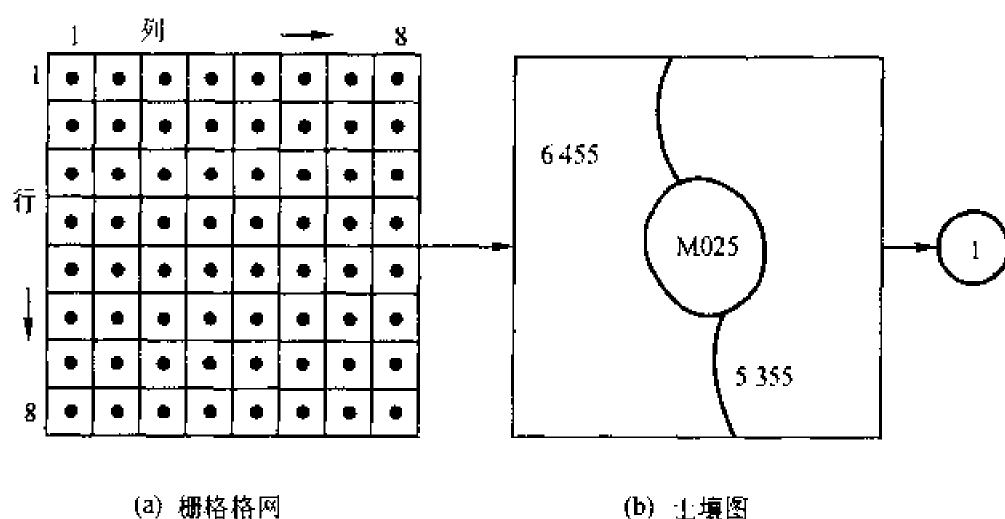
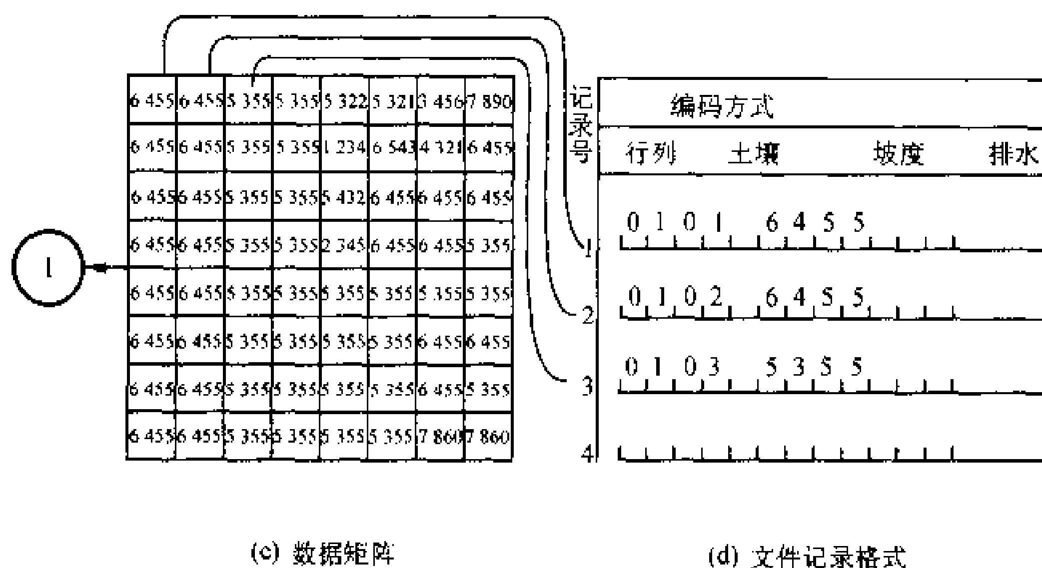


图 2－45　单位栅格长度占优法



(a) 栅格格网　　　　(b) 土壤图



(c) 数据矩阵　　　　(d) 文件记录格式

图 2－46　土壤图的栅格数据获取过程

(据 M.J.Rogoff，1982)

的过程。适合产生这种空间数据的图形数字化方法包括：透明格网采集输入、扫

描数字化输入及其他数据传输和转换输入等。

栅格数据的获取方法比较简单,以透明格网采集输入为例,具体步骤是,首先准备一张聚脂薄膜透明格网,格网的尺寸(边长)依要求而定;然后,将此格网叠置于原图上,根据单位格网交点归属法(图 2-43)、单位格网面积占优法(图 2-44)、单位格网长度占优法(图 2-45)等,可以直接获取相应的栅格数据。图 2-46为怎样获取土壤图栅格数据的一个实例。图 2-46(a)为格网,共有 8 行和 8 列,其中每个栅格单元都可以用行和列的坐标予以标识,例如,左上角一个栅格单元的坐标为(1,1),右下角一个栅格单元的坐标为(8,8)。图 2-46(b)为土壤图,表示出土壤类型的代码和分布界线。图 2-46(c)为获取的土壤图栅格矩阵,可以直接输入计算机。图 2-46(d)为土壤和其他有关要素的数据文件及其记录格式。

当区域范围不大,或者栅格单元的尺寸较大时,手工编码是简便可行的。但是,当区域范围较大,或者栅格单元的分辨率比较高,例如一幅 400×500 mm 的图幅,栅格单元的边长为 0.2 mm,则总图形数据量达到 5×10^6,这时靠手工编码非常费时,甚至不可能,需要采用数据类型转换方法,即由矢量数据向栅格数据的自动转换。其具体转换的算法参阅第 3 章 § 2。

栅格数据分为平面栅格数据和曲面栅格数据。以上讨论的是平面栅格数据,它对于全国或区域的土地资源和土地利用的分析、研究,以及与遥感数据的匹配应用等,都是非常方便的。若要获取按一定格网形式排列的地面点高程数据,或称为三维数据,则一般要经过数据取样、数据处理和数据记录三个过程,其具体方法详见第 3 章 § 5。

思考题

1. 通过实例说明 GIS 空间数据的基本特征及在计算机中的表示方法。

2. 简述空间数据的拓扑关系及其对 GIS 数据处理和空间分析有何重要意义?

3. 根据右转算法和左转算法的不同,写出由有向线段组成多边形区域的定义,并比较多边形边界与多边形区域的概念。

4. 试修改四叉树定义,使之适合存储数字化高程数据,并说明其存储结构的优缺点。

5. 叙述从某一空间地理对象到数据库或数据文件记录的演绎过程,并说明各个过程的作用。

6. 试写出图 2-47 的 DIME 数据文件和对多边形 113 进行联接编辑的算法步骤。比较多边形联接编辑与结点联接编辑的异同。

7. 如上图所示,设将该图纵边和横边 8 等分,

图 2-47

试建立:①该图层的栅格数据文件,并计算该栅格文件的数据冗余度;②该图层的游程编码数据文件(带索引文件);③该图层的线性四叉树数据文件及二维游程数据文件,并对数据压缩量作对比分析。

第3章 空间数据的处理

地理信息系统是采集、管理、处理、分析、建模和显示地理空间数据的数字系统,因此空间数据的处理是地理信息系统的重要功能之一。数据处理涉及的内容很广泛,主要取决于原始数据的特点和用户的具体需求,一般包括数据变换、数据重构、数据提取等内容。

数据变换指数据从一种数学状态到另一种数学状态的变换,包括几何纠正、投影转换和辐射纠正等,以解决空间数据的几何配准。数据重构指数据从一种格式到另一种格式的转换,包括结构转换、格式变换、类型替换等,以解决空间数据在结构、格式和类型上的统一,实现多源和异构数据的联接与融合。数据提取指对数据进行某种有条件的提取,包括类型提取、窗口提取、空间内插等,以解决不同用户对数据的特定需求。

数据处理是针对数据本身完成的操作,不涉及内容的分析。因此,空间数据的处理又称为数据形式的操作。以下分别介绍。

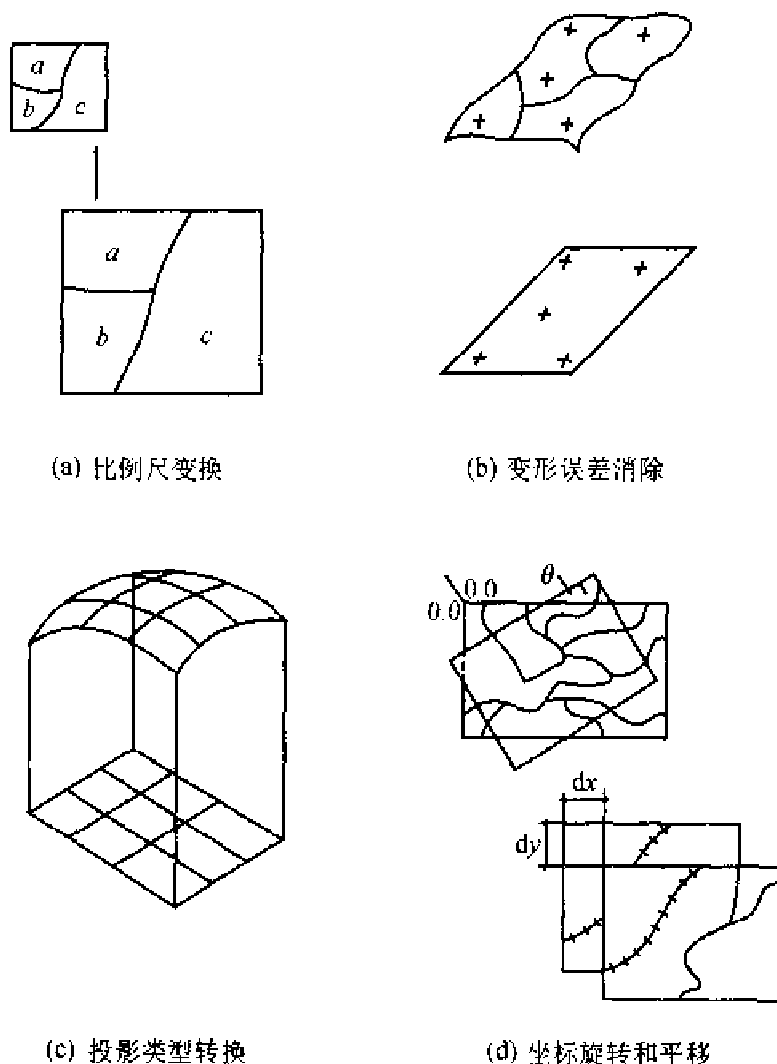
§1 空间数据的坐标变换

空间数据坐标变换的实质是建立两个平面点之间的一一对应关系,包括几何纠正和投影转换,它们是空间数据处理的基本内容之一。对于数字化地图数据,由于设备坐标系与用户确定的坐标系不一致,以及由于数字化原图图纸发生变形等原因,需要对数字化原图的数据进行坐标系转换和变形误差的消除。有时,不同来源的地图还存在地图投影与地图比例尺的差异,因此,还需要进行地图投影的转换和地图比例尺的统一(图 3-1)。

1.1 几何纠正

几何纠正是为了实现对数字化数据的坐标系转换和图纸变形误差的改正。现有的几种商业 GIS 软件一般都具有仿射变换、相似变换、二次变换等几何纠正功能。

仿射变换与相似变换相比较,前者是假设地图因变形而引起的实际比例尺在 x 和 y 方向上都不相同,因此,具有图纸变形的纠正功能。



(a) 比例尺变换 (b) 变形误差消除

(c) 投影类型转换 (d) 坐标旋转和平移

图 3-1 空间数据的坐标变换
(据 Jack Dangermond, 1984)

如图 3-2 所示，设 x, y 为数字化仪坐标，X、Y 为理论坐标，m_1、m_2 为地图横向和纵向的实际比例尺，两坐标系夹角为 α，数字化仪原点 O' 相对于理论坐标系原点平移了 a_0、b_0，则根据图形变换原理，得出坐标变换公式：

$$\begin{cases} X = a_0 + (m_1\cos\alpha)x + (m_2\sin\alpha)y \\ Y = b_0 - (m_1\cdot\sin\alpha)x + (m_2\cdot\cos\alpha)y \end{cases} \tag{3-1}$$

式中，设 $a_1 = m_1\cdot\cos\alpha$

$b_1 = -m_1\cdot\sin\alpha$

$a_2 = m_2\cdot\sin\alpha$

$b_2 = m_2\cdot\cos\alpha$

则式(3-1)可以简化为

$$\begin{cases} X = a_0 + a_1 x + a_2 y \\ Y = b_0 + b_1 x + b_2 y \end{cases} \quad (3-2)$$

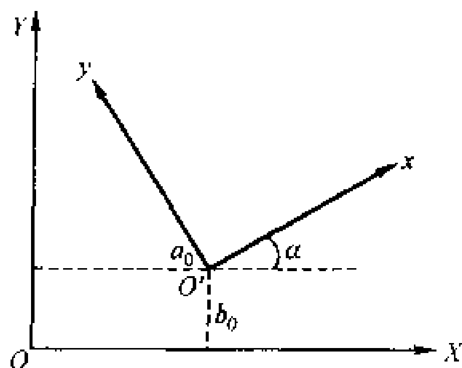


图3-2 坐标变换原理

式(3-2)含有6个参数 a_0、a_1、a_2、b_0、b_1、b_2，要实现仿射变换，需要知道不在同一直线上的3对控制点的数字化坐标及其理论值，才能求得上述6个待定参数。但在实际应用中，通常利用4个以上的点来进行几何纠正。下面按最小二乘法原理来求解待定参数：

设 Q_x、Q_y 表示转换坐标与理论坐标之差，则有

$$\begin{cases} Q_x = X - (a_0 + a_1 x + a_2 y) \\ Q_y = Y - (b_0 + b_1 x + b_2 y) \end{cases} \quad (3-3)$$

按照 $[Q_x^2] = \min$ 和 $[Q_y^2] = \min$ 的条件，可得到两组法方程：

$$\begin{cases} a_0 n + a_1 \sum x + a_2 \sum y = \sum X \\ a_0 \sum x + a_1 \sum x^2 + a_2 \sum x \cdot y = \sum x \cdot X \\ a_0 \sum y + a_1 \sum x \cdot y + a_2 \sum y^2 = \sum y \cdot X \end{cases} \quad (3-4)$$

和

$$\begin{cases} b_0 n + b_1 \sum x + b_2 \sum y = \sum Y \\ b_0 \sum x + b_1 \sum x^2 + b_2 \sum x \cdot y = \sum x \cdot Y \\ b_0 \sum y + b_1 \sum x \cdot y + b_2 \sum y^2 = \sum y \cdot Y \end{cases} \quad (3-5)$$

式中：n 为控制点个数；

x、y 为控制点的数字化坐标；

X、Y 为控制点的理论坐标。

由上述法方程，通过消元法，可求得仿射变换的待定参数 a_0、a_1、a_2 和 b_0、b_1、b_2。

仿射变换是GIS数据处理中使用最多的一种几何纠正方法。它的主要特性为：同时考虑到 x 和 y 方向上的变形，因此纠正后的坐标数据在不同方向上的长度比将发生变化。其他方法还有相似变换和二次变换等。

经过仿射变换的空间数据，其精度可用点位中误差表示，即

$$M_P = \pm\sqrt{\frac{[\Delta x^2 + \Delta y^2]}{n}} \quad (3-6)$$

式中：$\Delta x = X_{理论值} - X_{计算值}$；

$\Delta y = Y_{理论值} - Y_{计算值}$；

n 为数字化已知控制点的个数。

1.2 投影转换

当系统使用的数据取自不同地图投影的图幅时，需要将一种投影的数字化数据转换为所需要投影的坐标数据。投影转换的方法可以采用：

(1) 正解变换。通过建立一种投影变换为另一种投影的严密或近似的解析关系式，直接由一种投影的数字化坐标 x、y 变换到另一种投影的直角坐标 X、Y。

(2) 反解变换。即由一种投影的坐标反解出地理坐标（x、$y \rightarrow B$、L），然后将地理坐标代入另一种投影的坐标公式中（B、$L \rightarrow X$、Y），从而实现由一种投影的坐标到另一种投影坐标的变换（x、$y \rightarrow X$、Y）。

(3) 数值变换。根据两种投影在变换区内的若干同名数字化点，采用插值法，或有限差分法，或有限元法，或待定系数法等，从而实现由一种投影的坐标到另一种投影坐标的变换。

§2 空间数据结构的转换

如第 2 章表 2－7 所示，由于矢量数据结构和栅格数据结构各具有不同的优缺点，一般对它们的应用原则是：数据采集采用矢量数据结构，有利于保证空间实体的几何精度和拓扑特性的描述；而空间分析则主要采用栅格数据结构，有利于加快系统数据的运行速度和分析应用的进程。因此，在数据处理阶段，经常要进行两种数据结构的相互转换。

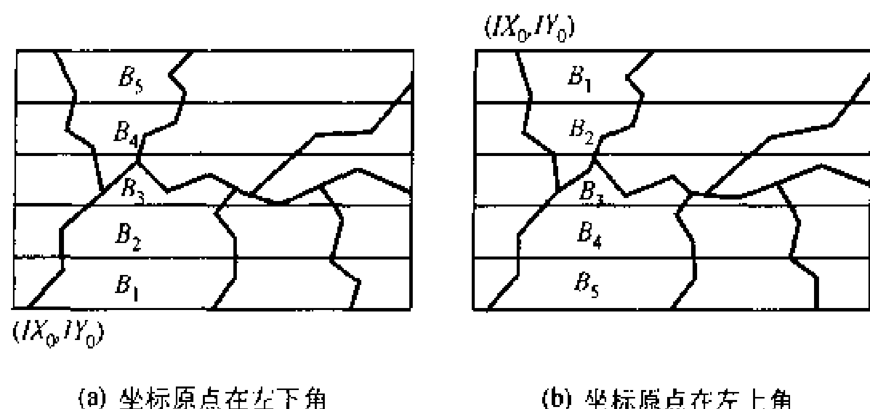
2.1 由矢量向栅格的转换

矢量向栅格转换处理的根本任务是通过一个有限的工作存储区，使得矢量和栅格数据之间不可避免的读写操作，限制在最短的时间范围内。根据转换处理时，基于弧段数据文件和多边形数据文件的不同，分别采用不同的算法。

1. 基于弧段数据的栅格化方法

弧段数据栅格化的算法可以分解为数据管理和转换计算两步。

(1) 数据管理。数据管理的任务是按照可用的工作存储区，将要建立的栅格数据体划分成数据段，对每个数据段或栅格带（图 3－3）建立相应的矢量数据文件。其方法是，对总体的原始矢量数据（弧段）文件进行一次性扫描，计算原点（IX_0，IY_0）、弧段总数 NA、栅格原始矢量数据（弧段）文件进行一次性扫描，计算原点（IX_0，IY_0）、弧段总数 NA、栅格行数 NR、列数 NC、分带数 NW、以及各个弧段按照它们在纵轴上的位置归入相应的子数据体，max（ID）表示该弧段归入的



(a) 坐标原点在左下角　　　　(b) 坐标原点在左上角

图3-3　栅格矩阵按不同坐标系分带示意图

栅格带最大的编号，min(ID)表示该弧段归入的栅格带最小的编号，ID 为弧段的识别码，D 为栅格的边长。这样，就实现了矢量数据分割，以便按照规划的栅格带，进行矢量向栅格的转换计算。

(2) 转换计算。转换计算的任务是将任意的 x、y 坐标转换为由行号(I)和列号(J)表示的栅格数据。具体转换的方法采用扫描线与有关的弧段求交，然后将交点的 x 坐标、该弧段对应的左右区码记入该扫描线所在行的数组，最后通过对一行所有 x 值按照由小至大排序、进行左右区码的配对和施行奇点的处理等技术，在相邻 x 值之间，逐段生成栅格数据，直到全部扫描行都完成从矢量向栅格的转换为止。

如图3-4所示，设 y_0 为通过某个栅格带中心的扫描线，该扫描线与弧段 C_1、C_2 的交点分别为(x_i, y_i)和(x_{i+1}, y_{i+1})；弧段 C_1 相邻两节点的坐标为(x_1, y_1)和(x_2, y_2)；弧段 C_2 相邻两节点的坐标为(x_1, y_1)和(x_3, y_3)。则根据两直线方程：

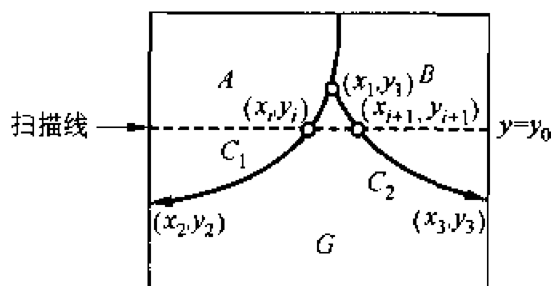


图3-4　扫描线与弧段相交示意图

$$\begin{cases} y_i = y_0 \\ \dfrac{y_i - y_1}{x_i - x_1} = \dfrac{y_2 - y_1}{x_2 - x_1} \end{cases} \tag{3-7}$$

和

$$\begin{cases} y_{i+1} = y_0 \\ \dfrac{y_{i+1} - y_1}{x_{i+1} - x_1} = \dfrac{y_3 - y_1}{x_3 - x_1} \end{cases} \tag{3-8}$$

可求得交点(x_i, y_i)和(x_{i+1}, y_{i+1})。将它们转换为栅格数据的行号(I)和列号(J)，对应于交点(x_i、y_i)的行、列号分别为：

$$\begin{cases} I_1 = \left[\frac{y_i}{D} \right] + 1 \\ J_1 = \left[\frac{X_i}{D} \right] + 1 \end{cases} \tag{3-9}$$

对应于交点(X_{i+1}、y_{i+1})的行、列号分别为：

$$\begin{cases} I_2 = \left[\frac{y_{i+1}}{D} \right] + 1 \\ J_2 = \left[\frac{x_{i+1}}{D} \right] + 1 \end{cases} \tag{3-10}$$

于是，可确定两交点间的栅格列数($J_2 - J_1$)及其属性值(G)。其算法过程用框图表示为图 3－5。

2. 基于多边形数据的栅格化方法

这种方法的原理是以多边形作为栅格化的处理单元，首先将一个多边形的 x 坐标数据，分别按照顺时针和逆时针的方向排序，即

$$x_1, x_2, \cdots, x_{n-1}, x_n$$

和

$$x_n, x_{n-1}, \cdots, x_2, x_1$$

对任一栅格点 P，设其坐标为 x_p、y_p，则根据该点的 x 值在上述两组坐标序列中相应区间出现的次数，可以求出相应个数的 y 值。设 y_a 对应于顺时针排列的 x 区间的内插值，y_b 对应于逆时针排列的 x 区间的内插值，则

$$y_a = y_i + \frac{y_i - y_{i-1}}{x_i - x_{i-1}} \cdot (x_p - x_{i-1})$$

$$y_b = y_j + \frac{y_j - y_{j+1}}{x_j - x_{j+1}} \cdot (x_p - x_{j-1})$$

设 A_1、A_2 和 A_3 为中间变量，而且 A_3 为 A_1 和 A_2 之和，并且 A_1 和 A_2 的取值由下式决定

$$A_1 = \begin{cases} 1 & \text{当 } y_a \geqslant y_p \text{ 时} \\ -1 & \text{当 } y_a < y_p \text{ 时} \end{cases}$$

$$A_2 = \begin{cases} 1 & \text{当 } y_b < y_p \text{ 时} \\ -1 & \text{当 } y_b \geqslant y_p \text{ 时} \end{cases}$$

则，如果 $|A_3| \geqslant 2$，该格网点 p 位于该多边形之内，将该多边形的属性值赋于该栅格。按照这种方法逐点处理，直到全部多边形生成栅格数据或栅格游程编码数据。

基于多边形数据的栅格化方法，还可以通过检验夹角之和和检验交点数的方法来实现。以下，讨论这两种方法：

(1) 检验夹角之和。如图 3－6 所示，设空间平面图形 *ABCDE* 在 xOy 投影面

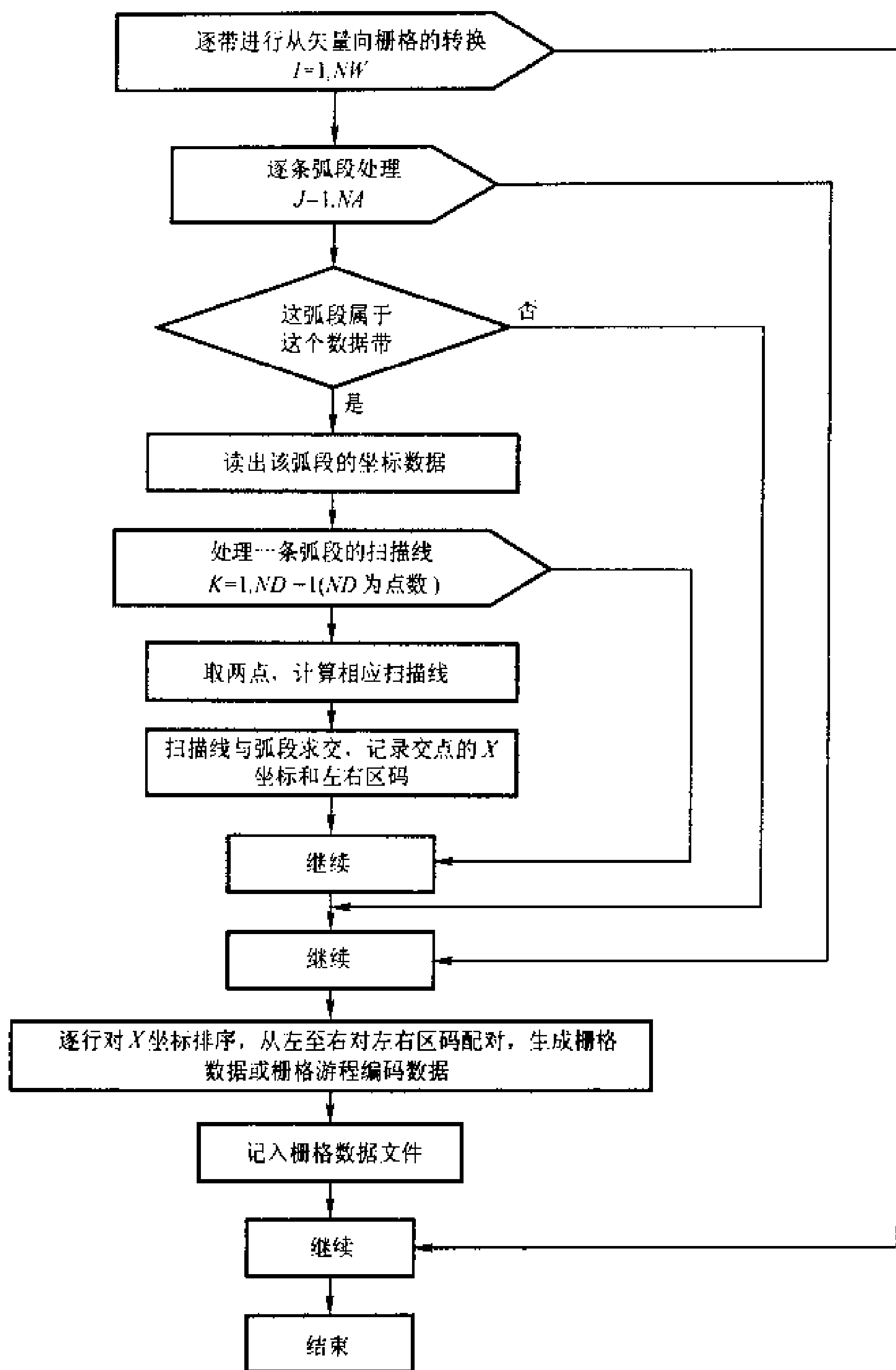


图3-5 基于弧段数据的栅格化转换算法

上的投影是 $A'B'C'D'E'$，空间另一点 P 在 xOy 投影面上的投影为 P'，P' 表示任一待判别的栅格点，依次将 P' 与 $A'B'C'D'E'$ 各点相连，令 α_i 分别为 $\angle A'P'B'$，

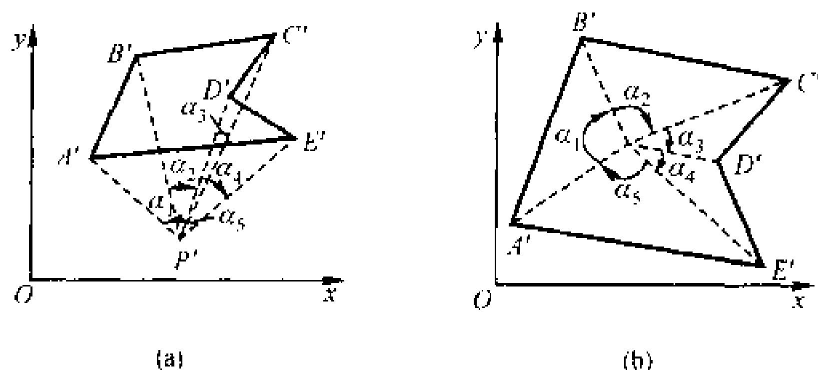


图 3-6 角度判断示意图

$\angle B'P'C'$, $\angle C'P'D'$, ……, 如果 $\sum_{i=1}^{5}\alpha_i = 0$, 则 P' 在多边形 $A'B'C'D'E'$ 之外(图 3-6(a)); 如果 $\sum_{i=1}^{5}\alpha_i = 2\pi$, 则 P' 在多边形 $A'B'C'D'E'$ 之内(图 3-6(b))。这里, α_i 的大小及方向,可由下式求得:

试以 $\angle A'P'B'$ 为例, $|\alpha_i|$ 可由余弦定理求出,

即
$$\overline{A'B'}^2 = \overline{A'P'}^2 + \overline{B'P'}^2 - 2|\overline{A'P'}|\cdot|\overline{B'P'}|\cdot\cos\alpha_1$$

$$\alpha_1 = \arccos\left(\frac{\overline{A'P'}^2 + \overline{B'P'}^2 - \overline{A'B'}^2}{2|\overline{A'P'}|\cdot|\overline{B'P'}|}\right) \tag{3-11}$$

而 α_i 的方向可以这样定义:

$$T_1 = \begin{vmatrix} x_{A'} - x_{P'} & y_{A'} - y_{P'} \\ x_{B'} - x_{P'} & y_{B'} - y_{P'} \end{vmatrix} \tag{3-12}$$

若 $T_1 < 0$, 则 α_1 为顺时针方向角;

若 $T_1 > 0$, 则 α_1 为逆时针方向角。

同理,由 T_i 的正、负,可以决定 α_i 的方向。

(2) 检验交点数。如图 3-7 所示,由任一待判别的栅格点 P' 向下作与 y 轴平行的射线,计算该射线与多边形 $A'B'C'D'E'$ 的交点数。若交点数为偶数,则栅

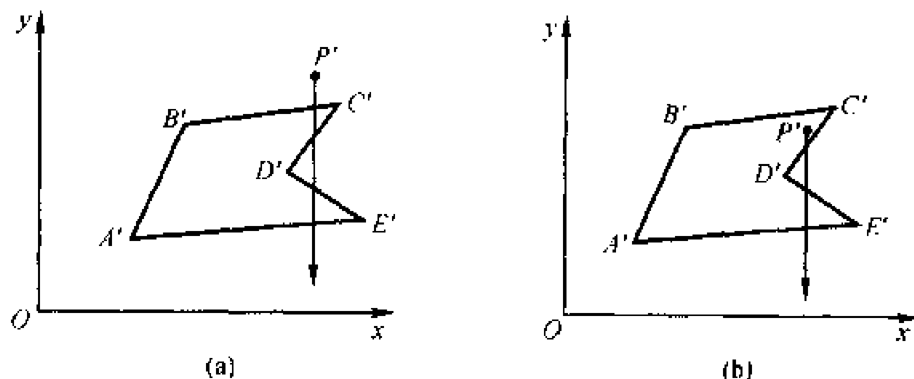


图 3-7 交点计数示意图

格点 P' 在多边形之外(图 3-7(a)),不予记录;若交点数为奇数,则栅格点 P' 在多边形之内(图 3-7(b)),予以记录,并将多边形的属性赋予该栅格点。重复上述过程,直至所有栅格单元填满该多边形为止。

以上方法统称为点与平面图形的包含性检验,或称为 point-in-polygon 分析。

2.2 由栅格向矢量的转换

栅格向矢量转换处理的目的,是为了将栅格数据分析的结果,通过矢量绘图装置输出,或者为了数据压缩的需要,将大量的面状栅格数据转换为由少量数据表示的多边形边界,但是主要目的是为了能将自动扫描仪获取的栅格数据加入矢量形式的数据库。转换处理时,基于图像数据文件和再生栅格数据文件的不同,分别采用不同的算法。

1. 基于图像数据的矢量化方法

图像数据是由不同灰阶的影像或线划,通过自动扫描仪(Scanner),按一定的分辨率进行扫描采样,得到以不同灰度值(0~255)表示的数据(图3-8)。目前

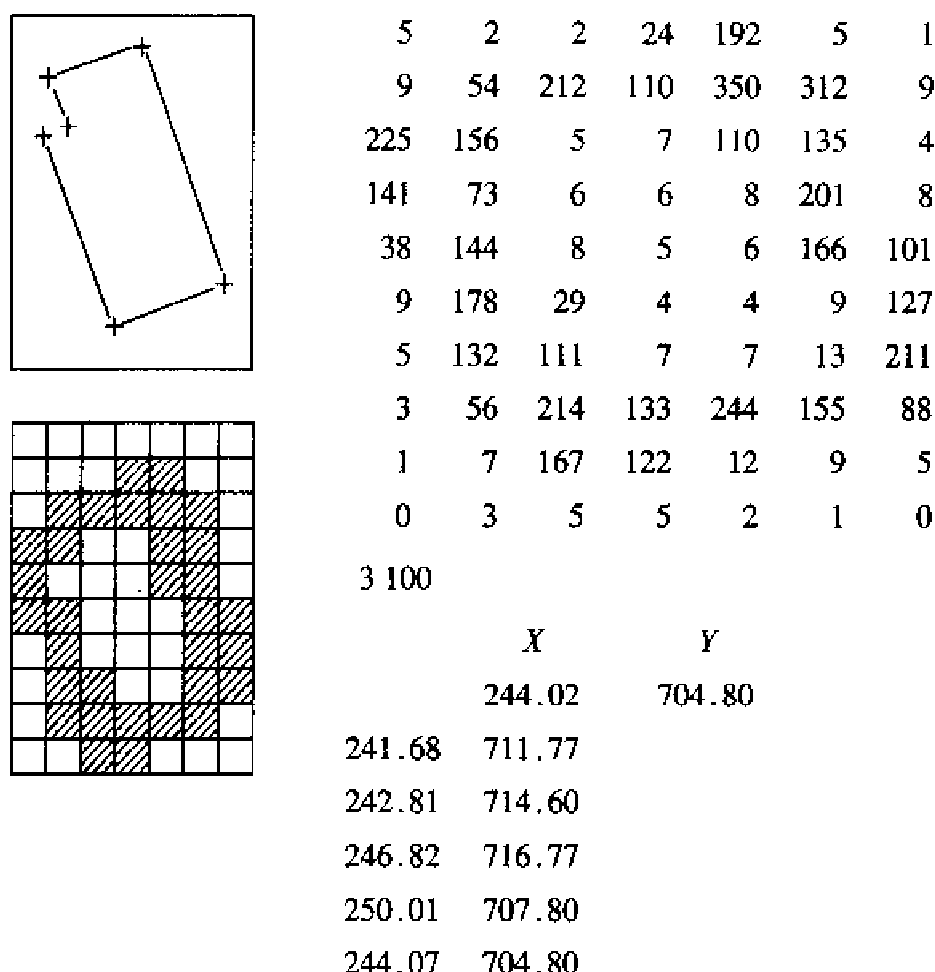


图 3-8 栅格与矢量数据表示法

(据 W. Linchtner 等,1987)

扫描仪的分辨率可达0.0125 mm,因此对一般粗度(例如0.1 mm)的线条,其横断面扫描后平均也有8个像元,而矢量化的要求只能允许横断面保持一个栅格的宽度,因此需要进行从栅格向矢量数据的转换。具体转换的步骤如图3-9所示。

(1) 二值化。线划图形扫描后产生栅格数据,这些数据是按从0~255的不同灰度值量度的(图3-9(a)),设以 $G(i,j)$ 表示,为了将这种256或128级不同的灰阶压缩到2个灰阶,即0和1两级,首先要在最大与最小灰阶之间定义一个阈值,设阈值为 T,则根据下式得到二值图(图3-9(b)):

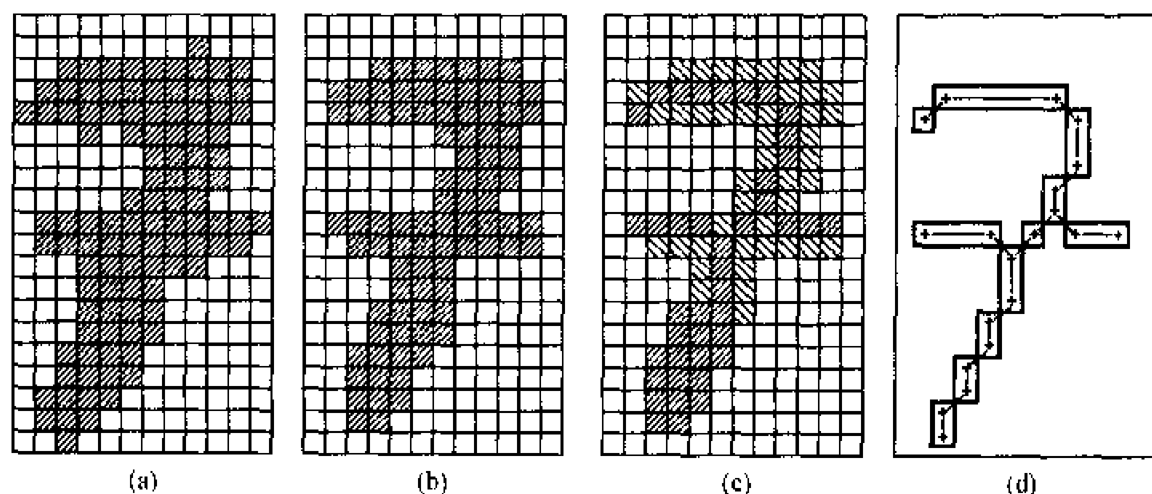


图3-9 栅格-矢量转换过程

(据W. Linchtner等,1987)

$$B(i,j)=\begin{cases}1 & \text{如果 } G(i,j)\geq T\\ 0 & \text{如果 } G(i,j)<T\end{cases}$$

(2) 细化。细化是消除线划横断面栅格数的差异,使得每一条线只保留代表其轴线或周围轮廓线(对面状符号言)位置的单个栅格的宽度。对于栅格线划的"细化"方法,可分为"剥皮法"和"骨架化"两大类。剥皮法的实质是从曲线的边缘开始,每次剥掉等于一个栅格宽的一层,直到最后留下彼此连通的由单个栅格点组成的图形。因为一条线在不同位置可能有不同的宽度,故在剥皮过程中必须注意一个条件,即不允许剥去会导致曲线不连通的栅格。这是这一方法的技术关键所在。其解决办法是,借助一个在计算机中存储着的,由待剥栅格为中心的3×3栅格组合图(图3-10)来决定。如图所示,一个3×3的栅格窗口,其中心栅格有八个邻域,因此组合图共有 2^8 种不同的排列格式,若将相对位置关系的差异只是转置90°、180°、270°或互为镜像反射的方法进行归并,则共有51种排列格式。显然,其中只有格式2、3、4、5、10、11、12、16、21、24、28、33、34、35、38、42、43、46和50,可以将中心点剥去。这样,通过最多核查256×8个栅格,便可确

定中间栅格点保留或删除，直到最后得到经细化处理后应予保留的栅格系列(图3-9(c))，并写入数据文件。

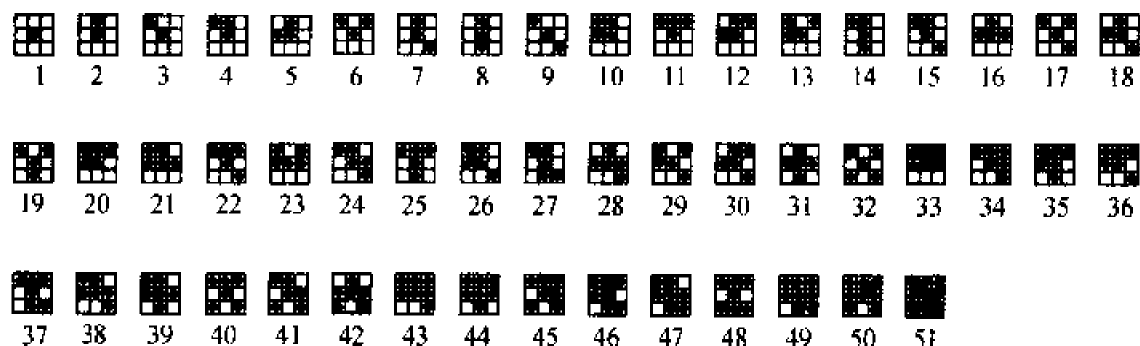


图3-10 3×3栅格组合图
(据W. Linchtner等,1987)

(3) 跟踪。跟踪的目的是将写入数据文件的细化处理后的栅格数据，整理为从节点出发的线段或闭合的线条，并以矢量形式存储于特征栅格点中心的坐标(图3-9(d))。跟踪时，从图幅西北角开始，按顺时针或逆时针方向，从起始点开始，根据八个邻域进行搜索，依次跟踪相邻点。并记录节点坐标，然后搜索闭曲线，直到完成全部栅格数据的矢量化，写入矢量数据库。

2. 基于再生栅格数据的矢量化方法

再生栅格数据是根据弧段数据或多边形数据生成的栅格数据。这种数据除为了与图像数据相匹配，需要加入数据库，一般地，只提供分析应用，不需作为永久文件保存。而作为永久文件保存的是原始的矢量数据文件，包括结点坐标文件、弧段文件、多边形文件及多边形内部点文件。这种再生栅格数据的矢量化，其主要目的是为了通过矢量绘图装置输出，具体的矢量化算法描述如下：

具体处理时，首先，对栅格数据按行扫描，找出位于各类型边界的栅格单元，并将边界内部具有相同值或同质的栅格单元以一种显著不同的符号进行充填，产生只记录类型边界栅格值的文件；其次，建立对类型边界栅格单元的追踪算法，寻找同质区的闭合界线，同时计算其坐标，并整理成有序(按顺时针或逆时针方向)的坐标数组；最后，处理相邻类型的公共边界，将按区域单元建立的数据结构转换为按线段链建立的数据结构，以便实现任意区域或类型数据的提取、综合、分析和制图输出。

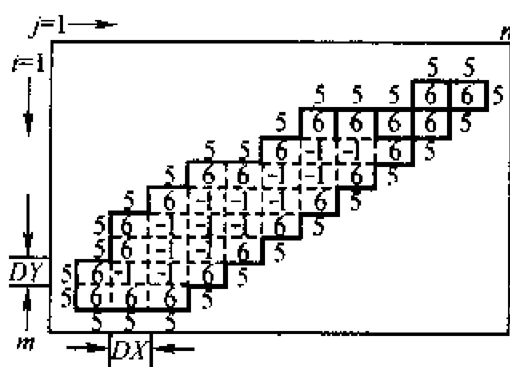


图3-11 将类型边界内部的同质栅格用(-1)充填

如图 3-11 所示，设栅格数据的坐标原点位于图幅的左上角，并且 $i=1,2,3,\cdots,m$；$j=1,2,\cdots,n$。当在栅格数据中，将类型边界内部的同质栅格单元用一种显著不同的符号（例如，-1）充填时，并设 $IXY(i,j)$ 为栅格数值，则栅格与栅格之间存在三种不同性质的交界线，即外边界（以粗线表示）、公共边界（以细线表示）和内边界（以虚线表示）。如果栅格单元四角点的坐标 $X(i_1,i_2)$、$Y(i_1,i_2)$ 表示，其中 i_1 为栅格单元 $(i、j)$ 四条边的识别码 $(i_1=1,2,3,4)$；i_2 表示其中每条边起讫点的识别码 $(i_2=1,2)$。则凡与某一类型的值相同的栅格单元 (i,j) 四角点的坐标（图 3-12），可用下式进行计算：

$$X(i_1,i_2)=(j-1)*DX \text{ 和 } j*DX$$

$$Y(i_1,i_2)=(j-1)*DY \text{ 和 } i*DY$$

式中：DX,DY 为栅格单元的边长；

$i=2,3,\cdots,m$；

$j=2,3,\cdots,n$。

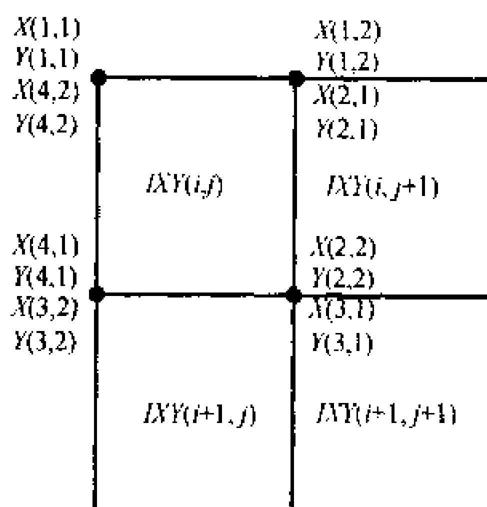


图 3-12 栅格单元四边和四端点的编号

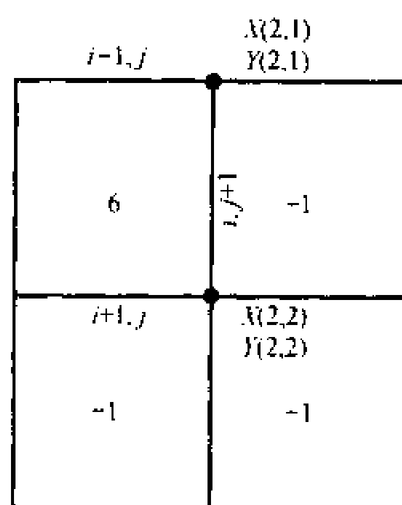


图 3-13 栅格和单元 (i,j) 的第二边是内边界

然后按照下述步骤来建立外边界一组有序的坐标，以供绘制类型图的边界线：

(1) 识别内边界，并将内边界端点的坐标置零。设 i,j 为栅格单元的序号，则它的四边和相邻栅格单元的序号如图 3-13 所示。显然，判别每个栅格单元的四边是否为内边界，也就是确定它的相邻四个栅格单元的值 $IXY(i,j)$ 是否小于零。例如，以栅格单元 (i,j) 的第二边为例，因为 $IXY(i,j+1)<0$，所以该边为内边界，并将其坐标置零，即

$$X(2,1)=0, \quad Y(2,1)=0$$

和

$$X(2,2)=0,\quad Y(2,2)=0$$

直到找出同质栅格单元的全部内边界,并将其坐标置零。

(2) 识别公共边界,并将公共边界端点的坐标置零。由于在计算每个栅格

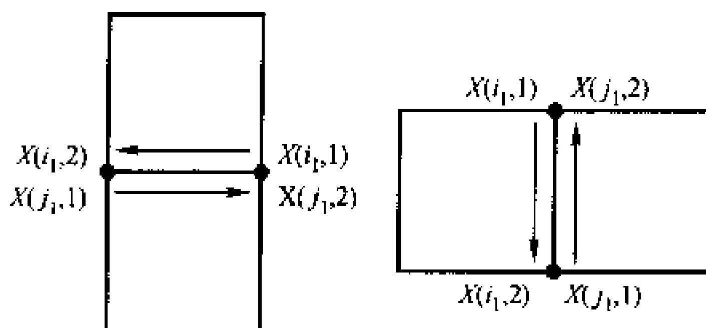


图 3-14 相邻栅格之间的公共边界

单元四角点的坐标时,各个栅格的相邻边都有两次重复(图 3-14),因此可以利用同一边两端点坐标值重复的条件来判定公共边界。设 i_1、j_1 表示同一边在不同栅格中的编号,其中 i_1 表示前一栅格边的编号,j_1 表示后一栅格边的编号,则识别公共边的条件是要同时满足下列四个等式:

$$X(i_1,2)=X(j_1,1)$$

$$Y(i_1,2)=Y(j_1,1)$$

$$X(i_1,1)=X(j_1,2)$$

$$Y(i_1,1)=Y(j_1,2)$$

一旦这个条件成立,就将相应边两端点的坐标置零,直至找出全部公共边界。

(3) 追踪外边界,并将外边界各点的坐标按首尾连接的顺序排列。当同质栅格之间的上述三种不同性质的交界线,经过内边界与公共边界置零处理以后,剩下凡非零的端点坐标即为外边界。但是要提供存储和绘图,必须将这些坐标按首尾连接的顺序排列。为此,要采用逐点追踪的方法(图 3-15),其算法如下:

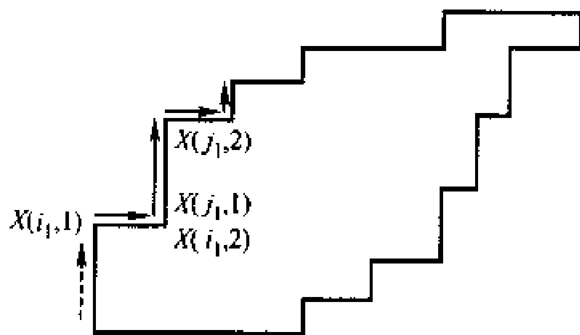


图 3-15 逐点追踪外边界的方法

图 3－16　原始栅格数据

① 如果 $X(i_1,1)\neq 0$ 和 $X(i_1,2)\neq 0$，则找到外边界的一段起始边，及其端点坐标 $X(i_1,1)$、$Y(i_1,1)$和 $X(i_1,2)$、$Y(i_1,2)$。

② 如果 $X(i_1,2)=X(j_1,1)$和 $Y(i_1,2)=Y(j_1,1)$，则找到相邻点及其坐标 $X(j_1,2)$、$Y(j_1,2)$。

③ 按照第②步的同样方法，继续追踪，直至又回到 $X(i_1,1)$、$Y(i_1,1)$点，表示类型的外边界线自行闭合，得到按同质类型建立的一组边界线坐标数据。但是，这种数据在相邻类型的边界上是重复的，需要进行处理、删去相邻边界线重复的数据，以免公共边界的两次重复绘图。需要时，还可以作边界线的光滑处理。

(4) 建立绘图数据文件。通过公共边界重复数据的删除，最后按照：

邻域 a_{i_1}, a_{i_2}	线段链点数 n_i	坐标 $x_{i_k}, y_{i_k}, k=1,\cdots,n_i$

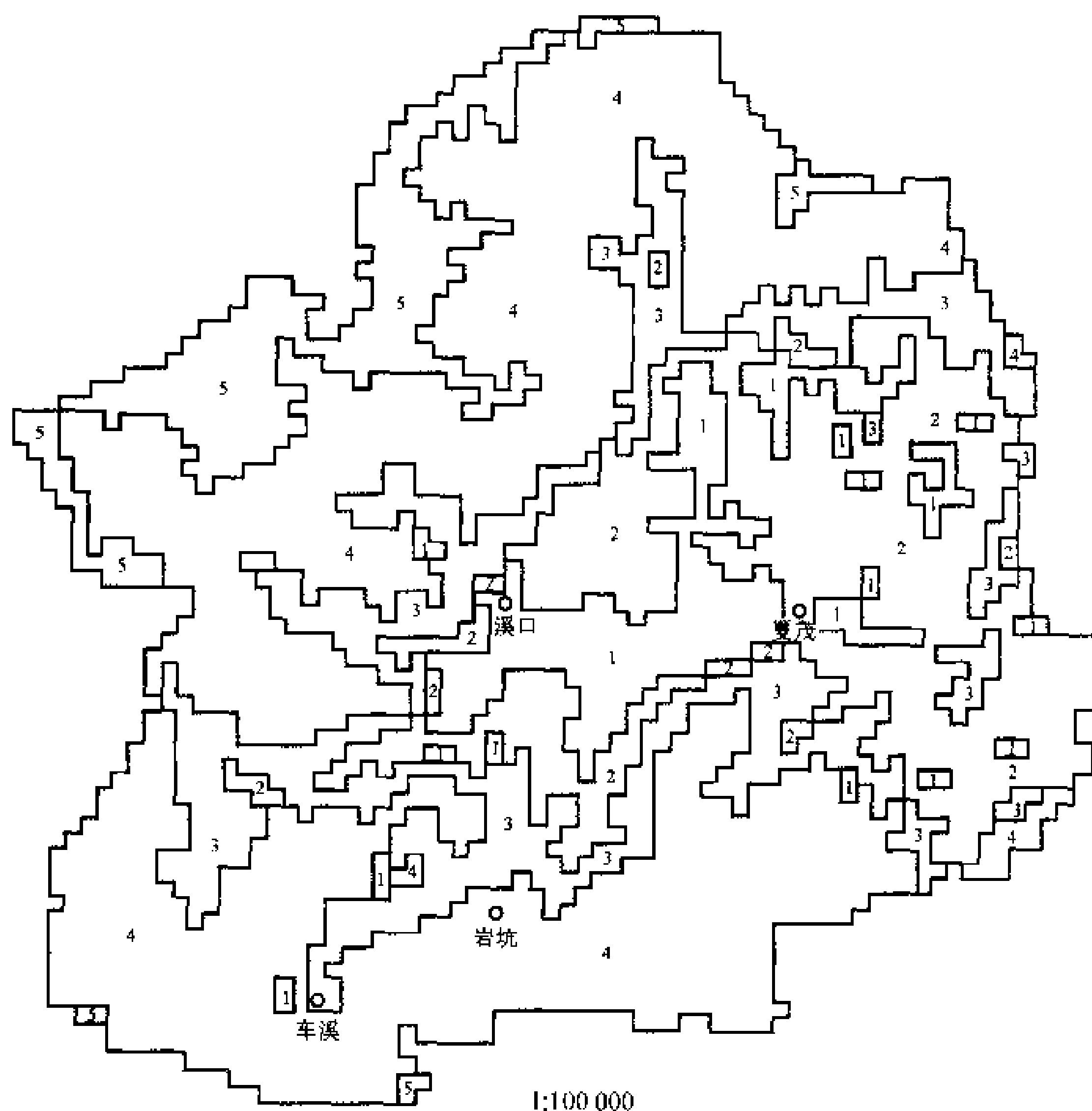


图 3-17 根据栅格数据自动绘制的地表类型图

1. 平地 2. 岗丘 3. 丘陵 4. 低山 5. 中山

的数据结构进行存储,以建立供绘图和分析用的矢量数据文件。图 3-17 即根据原始栅格数据(图 3-16)输出的矢量图。

§3 多源空间数据的融合

GIS 技术经过近 40 年的发展和应用,已经积累了大量的数据资源。但是,由于地理数据的多语义性、多时空性、多尺度性、获取手段的多样性、存储格式的不同以及数据模型与数据结构的差异等,导致多源数据的产生,给数据的集成和信息共享带来困难。为了实现空间数据的共享,特别是随着因特网的发展、数字地球的兴起和 GIS 应用的日益深入,多源数据的融合已成为 GIS 设计者和用户的共同要求。

3.1 遥感与 GIS 数据的融合

众所周知,借助遥感技术获得的信息具有周期动态性、信息丰富、获取效率高等优势,而 GIS 则具有高效的空间数据管理和灵活的空间数据综合分析能力。两者的数据融合,有利于增强多重数据的复合能力,改善遥感信息提取的及时性和可靠性,便于利用遥感影像辅助 GIS 空间数据的获取与更新,有效地提高各类数据的使用率。遥感与 GIS 数据的融合,目前最常用的方法具体表现为:

(1) 遥感图像与图形的融合。经过正射纠正后的遥感影像,与数字地图信息融合,可产生影像地图。这种影像地图具有一定的数学基础,有丰富的光谱信息与几何信息,又有行政界线和属性信息,直接提高了用户的可视化效果。

(2) 遥感数据与 DEM 的融合。DEM 代表精确的地形信息,它与遥感数据的融合,有助于实施遥感影像的几何校正与配准,消除遥感图像中因地形起伏所造成的像元位移,提高遥感图像的定位精度,同时 DEM 可参与遥感图像的分类,改善分类精度。

(3) 遥感图像与地图扫描图像的融合。将地图扫描图像与遥感图像配准叠合,可以从遥感图像中快速发现已发生变化的区域,进而实现 GIS 数据库的自动/半自动快速更新。

3.2 不同格式数据的融合

由于 GIS 软件的多样性,每种 GIS 软件都有自己特定的数据模型,造成数据存储格式和结构的不同。例如,ARC/INFO 是一种基于显式地描述空间地理实体的全拓扑关系的地理信息系统,各种要素的拓扑关系,如节点与弧的连接关系、面与面的邻接关系等都显示地存放在相应的属性表上。而 MapInfo 则是以实体

结构为基础的地理信息系统,在它的各种要素属性表上并不存放任何拓扑关系信息。因此,要实现两种系统之间数据的连接,首先必须解决不同系统之间空间数据模型的转换。

目前不同 GIS 软件系统使用的空间数据格式主要有:ESRI 公司的 ARC/INFO Coverage、ArcShape Files、E00 格式;Autodesk 公司的 DXF 和 DWG 格式;MapInfo 公司的 MIF 格式;Intergraph 公司的 DGN 格式等等。解决这些不同格式数据之间的融合,主要有以下几种方法:

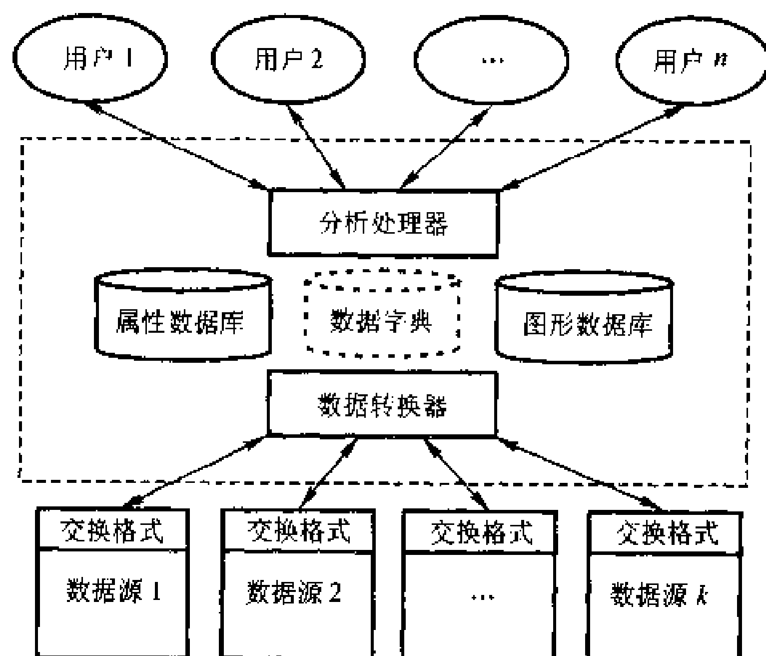


图3-18 单一数据库集成多源数据模式

(1) 基于转换器的数据融合。在这种模式下,数据转换一般通过交换格式进行(图3-18)。例如,要转换 MapInfo 的 Tab 文件到 ARC/INFO 的 Coverage,首先需要使用 MapInfo 软件将 Tab 文件输出为 E00 或 DXF 文件,然后运行 ARC/INFO 将 E00 或 DXF 文件转换为 ARC/INFO Coverage。这是目前 GIS 系统数据融合的主要方法。其存在的主要问题是:数据转换过程复杂,转换次数频繁,系统内部的数据格式需要公开,但转换采用的技术不公开等。

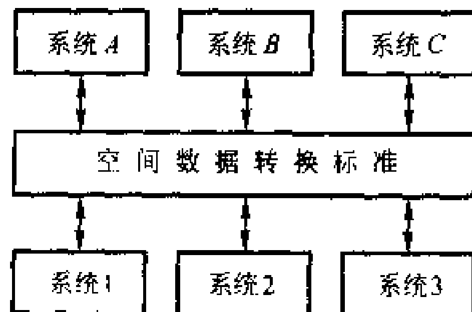


图3-19 基于数据转换标准的数据融合

(2) 基于数据标准的数据融合。这种方法是采用一种空间数据的转换标准来实现多源 GIS 数据的融合(图3-19)。例如,美国国家空间数据协会(NSDI)制定了统一的

空间数据格式规范 SDTS(spatial data transformation standard),包括几何坐标、投影、拓扑关系、属性数据、数据字典,也包括栅格和矢量等不同空间数据格式的转换标准。根据 SDTS,目前许多 GIS 软件提供了标准的空间数据交换格式,例如 ARC/INFO 的 SDTSIMPORT 和 SDTSEXPORT 模块等,可供其他系统调用。这种转换方法能处理多个数据集,转换次数少,系统内部的数据格式不需公开,但转换采用的技术需要公开等。

(3) 基于公共接口的数据融合。基于公共接口的数据融合模式又称为数据互操作模式。接口相当于一种规程,它是大家都遵守并达成统一的标准。在接口中不仅要考虑数据格式和数据处理,而且还要提供对数据处理应采用的协议,各个系统通过公共接口相互联系,而且允许各自系统内部数据结构和数据处理各不相同,具体如图 3-20 所示。例如,OGC(Open GIS Consortium)为数据互操作制定了统一的规范,从而使一个系统同时支持不同的空间数据格式成为可能。根据 OGC 颁布的规范,可以将提供数据源的软件称为数据服务器(Data Servers),将使用数据的软件称为数据客户(Data Clients),数据客户使用某种数据的过程就是发出数据请求,由数据服务器提供服务的过程,其最终目的是使数据客户能读取任意数据服务器提供的空间数据。这种基于接口的 OGC 规范正得到 OMG 和 ISO 的承认,将成为一种国际标准。其主要特点是独立于具体平台,转换技术高度抽象,数据格式不需公开,代表着数据共享技术的发展方向。

(4) 基于直接访问的数据融合。直接数据访问指在一个 GIS 软件中实现对其他软件数据格式的直接访问,用户可以使用单个 GIS 软件存取多种数据格式。直接数据访问不仅避免了繁琐的数据转换,而且在一个 GIS 软件中访问某种软件的数据格式,不要求用户拥有该数据格式的宿主软件,更不需要该软件的运行,这为多源数据的融合提供了更为实用便捷的支持。例如,Intergraph 公司推出的 Geomedia 系列软件提供了这种支持(图 3-21)。

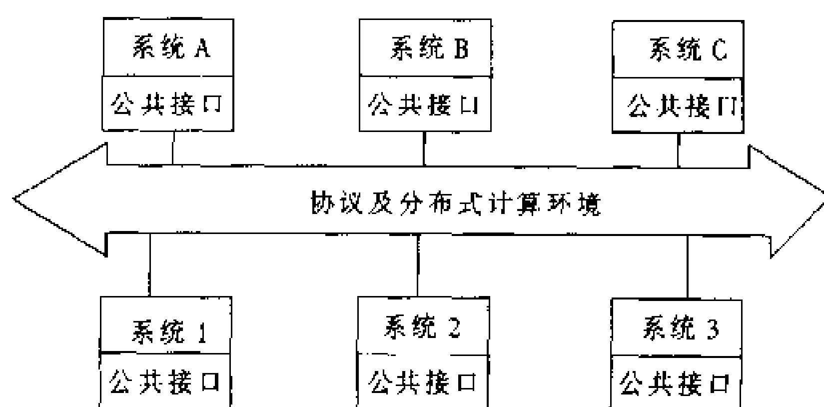


图 3-20 GIS 互操作示意图

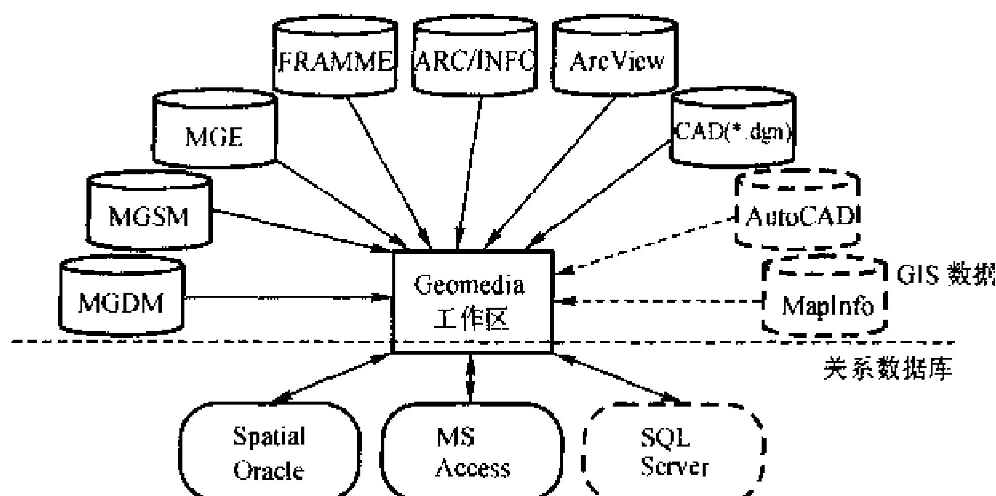


图 3-21　Geomedia 支持多源数据示意图

§4　空间数据的压缩与综合

4.1　空间数据的压缩

1. 数据压缩的意义

将数据组织为数据库，是建立地理信息系统的一个重要环节。利用现代的数据采集系统来量化空间要素，其分辨率目前可达 5 μm 的数量级，对如此庞大的坐标数据量，不采取数据压缩技术，整个系统在存储空间和处理时间上都将承受巨大的压力，而导致系统的失败。此外，随着处理空间数据的比例尺发生变化，同样也存在数据压缩的需要。

所谓数据压缩，即从所取得的数据集合 S 中抽出一个子集 A，这个子集作为一个新的信息源，在规定的精度范围内最好地逼近原集合，而又取得尽可能大的压缩比。压缩比表示曲线信息载量减少的程度。设曲线的原点序 A：

$$\{A_1, A_2, \cdots, A_n\}$$

通过数据压缩处理后，获得新子序 A'

$$\{A_{s_1}, A_{s_2}, \cdots, A_{s_n}\}$$

则压缩比 a 为

$$a = \frac{m}{n} \geqslant 1$$

显然，a 值的大小，既与曲线的复杂程度、缩小倍数、精度要求、数字化取点的密度等因素有关，又与数据压缩技术本身有关(例如中间点递推与否等)。

2. 曲线上点的压缩

设曲线由点序$\{P_1, P_2, \cdots, P_R\}$构成，则给定反映其坐标值的两数组

$$(x_1, x_2, \cdots, x_R)\text{和}(y_1, y_2, \cdots, y_R)$$

在自动抽取特征点时，处理区间由 P_M(起点)到 P_N(终点)。根据一条曲线的起点和终点建立直线方程(图 3-22)：

$$\frac{y - y_M}{x - x_M} = \frac{y_M - y_N}{x_M - x_N}$$

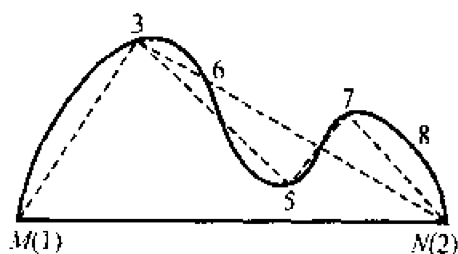


图 3-22 曲线上点的筛选

化成一般式 $Ax + By + C = 0$

其中

$$A = \frac{y_M - y_N}{D}$$

$$B = \frac{x_M - x_N}{D}$$

$$C = \frac{x_M y_N - x_N y_M}{D}$$

$$D = \sqrt{(y_M - y_N)^2 + (x_M - x_N)^2}$$

若 P_i 为 P_M 到 P_N 间任一点，

则有 $$d_i = |Ax_i + By_i + C|$$

为 P_i 到直线 $P_M P_N$ 的距离。

取 $d_k = \max(d_M, \cdots, d_N)$(约定等值时取第一个)，并给开关量 P 赋值

$$P = \frac{1}{2}\mathrm{sign}(d_k - \varepsilon)[1 + \mathrm{sign}(d_k - \varepsilon)]$$

即

$$P = \begin{cases} 1, & \text{当 } d_k > \varepsilon \text{ 时} \\ 0, & \text{当 } d_k \leqslant \varepsilon \text{ 时} \end{cases}$$

式中，ε 为控制数据压缩的极差(被舍去点距离取点连线之间的最大偏差，一般取为 < 0.2 mm)。当 $P = 0$ 时，P_N 作为留取点抽出，并依次排在前一个留取点之后。

这样，M、N 的初始值是 1 和 R，以后，当某段曲线不能以直线逼近时，则进一步处理从原起点到偏差最大点之间的曲线段；反之，则再处理原终点到距其最近一个被记录点间的一段。直至多余点全部被删除，以实现曲线信息量的压缩。这种数据压缩方法又称为特征点筛选法。它是以信息丢失为代价，换取空间数据容量的缩小。

3. 面域栅格数据的压缩

栅格数据是地理信息系统另一种常用的数据类型，扫描数字化数据、矢量－栅格转换后的数据、遥感数据以及 DTM 数据等，都属于这种数据类型。如前所述，栅格数据的分辨率与栅格数据量之间，呈平方指数率的函数关系。例如，一

幅 1m×1m 的地图,若采用 1 mm 的栅格尺寸,就有 10^6 个栅格数据,而一幅陆地卫星 MSS 图像有 $3\,240\times2\,340\approx7.58\times10^6$ 个像元。对如此庞大的数据量,不仅要考虑数据存储容量的压缩,而且还要顾及地理信息系统对数据访问的效率。常用的栅格数据压缩方法已在第 2 章 § 3 作了详细的论述。

4.2 空间数据的综合

存储在信息系统数据库中的数据,是提供长期使用和为多种目的服务的。当需要进行专门的数据分析时,常常需要先对从数据库中提取的数据作定向处理,这些定向处理包括数据属性的重新分类、空间图形的化简和图形特征的内插,以构成数据新的使用形式。

面域邻接线段的删除,是由于数据属性的重新分类和空间图形的化简而伴随需要的一种数据压缩形式,它包括相邻界线的删除和共同属性的合并,其过程如图 3-23 所示。

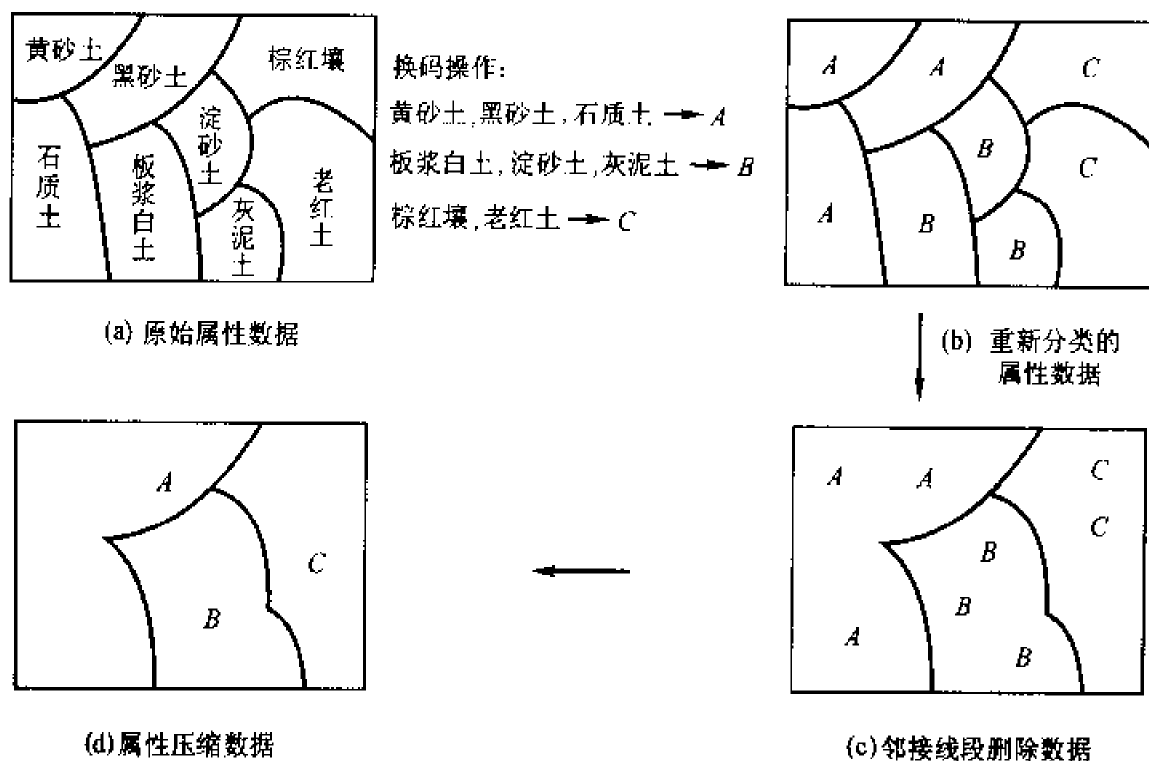


图 3-23 数据综合的过程

因此,当对属性性质相似的相邻两个面域合成一类相同的代码,或者当从多重属性的数据库中,只需要建立单要素的数据文件时,常需要采用这种压缩法来处理数据。否则,图 3-23 的输出形式将如图 3-24 所表现的形式。显然,图 3-24 不仅数据冗余度大,而且影响空间数据比例尺的进一步缩小。

面域之间界线的自动删除,可以通过构成每一面域的线段链,删去其中共同

的线段，然后重新建立合并多边形的线段链表。其算法过程如图 3 - 25 所示。

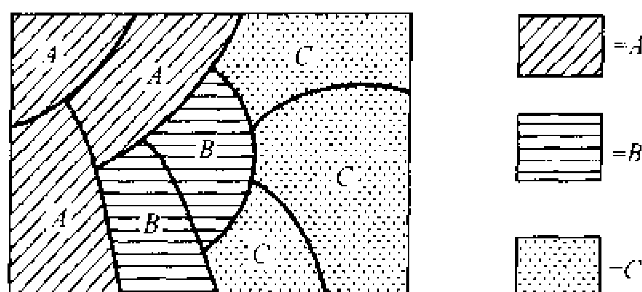


图 3 - 24　用晕线符号表示数据的综合

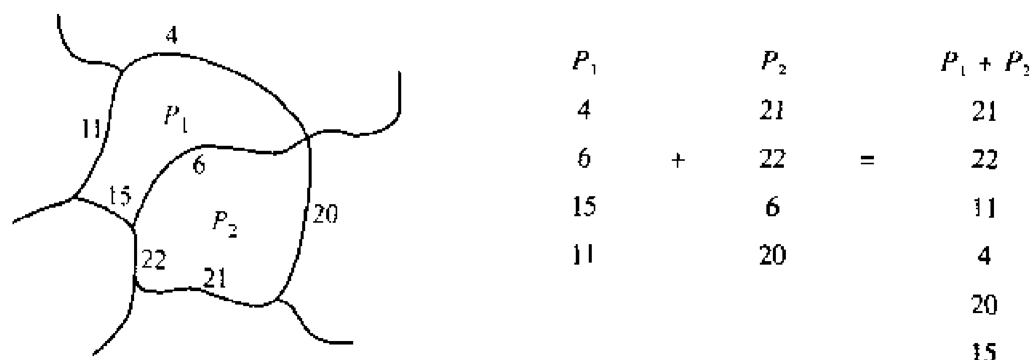


图 3 - 25　面域公共边界线的自动删除

§5　空间数据的内插方法

空间数据的内插可以作如下简单的描述：设已知一组空间数据，它们可以是离散点的形式，也可以是分区数据的形式，现在要从这些数据中找到一个函数关系式，使该关系式最好地逼近这些已知的空间数据，并能根据该函数关系式推求出区域范围内其他任意点或任意分区的值。这种通过已知点或分区的数据，推求任意点或分区数据的方法就称为空间数据的内插。根据已知点和已知分区数据的不同，将空间数据内插分为点的内插和区域的内插。它们是地理信息系统数据处理常用的方法之一，广泛用于等值线自动制图、数字高程模型的建立、不同区域界线现象的相关分析和比较研究等等。

5.1　点的内插

点的内插是研究具有连续变化特征现象（例如地面高程等）的数值内插方法。根据内插精度的不同，可以将点内插法分为精确和概略的两类（图 3 - 26）。但是，不论采用何种内插方法，建立按一定网络（例如三角网、格网等）形式排列的地面点高程数据，一般需要经过数据取样，数据处理和数据记录三个过程。数

据取样是指数据点的选取和坐标的确定；数据处理是以数据点作为控制基础，用某一数学模型来模拟地表面，进行内插加密计算，确定三角网或格网节点处的特征值；数据记录是将建立的节点特征值记录于存储器内，以供分析应用。

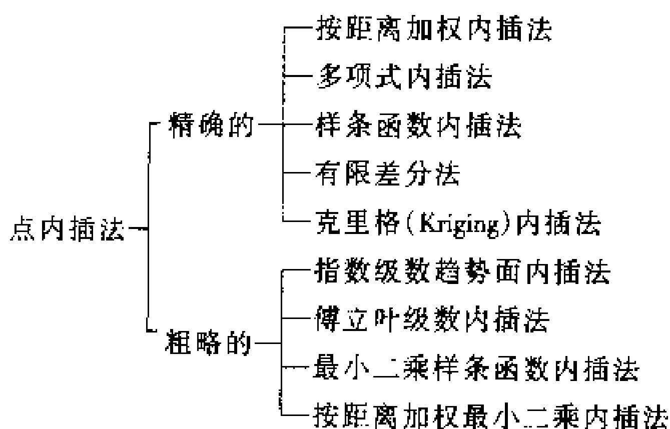


图3-26 点内插法的分类

1. 数据取样

建立按一定格网形式排列的地面点高程，一般先从现有地图的等高线上进行数据取样。取样点可以按地性线(山脊线、山谷线、坡度变换线)，沿等高线或沿断面线布设，将数据点选择在地性线坡度改变处，或沿等高线在方向改变的地点，即根据地形变化取点。这样，数据点落在地形特征点上，能很好地控制地表

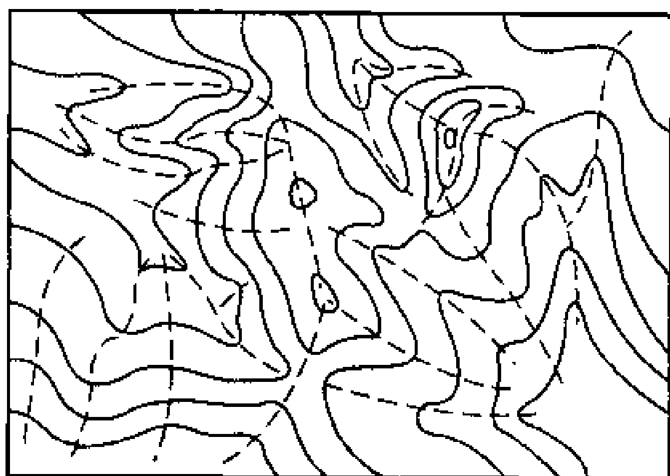


图3-27 沿等高线的地性线布设取样点

面形态(图3-27)，称为随机取样方案。该方案任意两点间直线部分相对于地面高程上的最大差距，应按系统的功能、地面形态特征和对地形分析的精度要求来确定，以保证具有足够的地面信息，作为数据处理的控制依据。具体取样时，利用数字化仪获取等高线上具有特征意义的各点的 x、y、z 数据，并分别建立包含这三种数据的高等线目标文件及每条等高线的索引文件。以下是等高线数据取样索引文件的一个具体例子。

LL	LG	LC	LP	IX_1	IY_1	IX_2	IY_2
20	1	1	20	785	371	895	370
25	1	21	69	643	489	981	370
30	1	90	88	641	567	1 013	370
35	1	178	159	641	610	1 014	371
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
80	4	384	6	2 058	862	2 040	824

LL　等高线高程(Z)值

LG　等高线目标文件的区号(每区由 1 000 个记录组成)

LC　每条等高线在目标文件中的地址(记录号)

LP　该条等高线的取样点数

IX_1, IY_1　该条等高线第一个取样点的坐标值

IX_2, IY_2　该条等高线最后一个取样点的坐标值

所以,由索引文件可以从目标文件中检索到每条等高线的起始记录号 R_1 和终止记录号 R_2,即

$$R_1 = (LG - 1) * 1\,000 + LC$$
$$R_2 = (LG - 1) * 1\,000 + LC + LP - 1$$

从而可以找到任意等高线上各点的 x、y、z 数据。

当按一定间隔的格网取样时,取一个二次曲线来代表地表面相应部分的曲线,而用间隔中点的线性内插值与二次曲线内插值所得到的断面高程的差异来判断格网密度是否恰当。如图 3-28 所示,设拟定的取样点 1、2 和 3,相应的高程为 h_1、h_2 和 h_3,取等距离间隔为单位数,则间隔中点的高程为:

线性内插值 $h' = \frac{1}{2}(h_1 + h_2)$

二次曲线内插值 $h'' = \frac{1}{8}(6h_2 + 3h_1 - h_3)$

其高程差异 $\delta_h = \frac{1}{8}(2h_2 - h_1 - h_3)$

此值应在数字地面模型精度要求的限差之内,否则要缩短格网间距。对地形变化显著的地方,要辅以地形特征点为数据点,以提高数据点控制地面形态特征的能力。

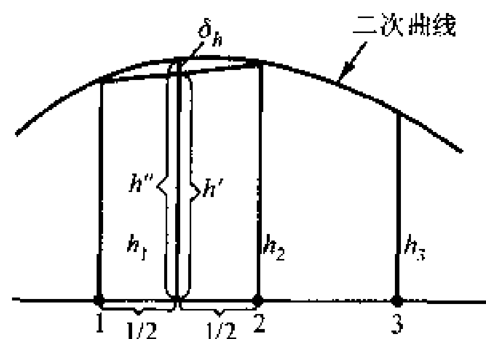


图 3-28 高程内插间距的确定方法

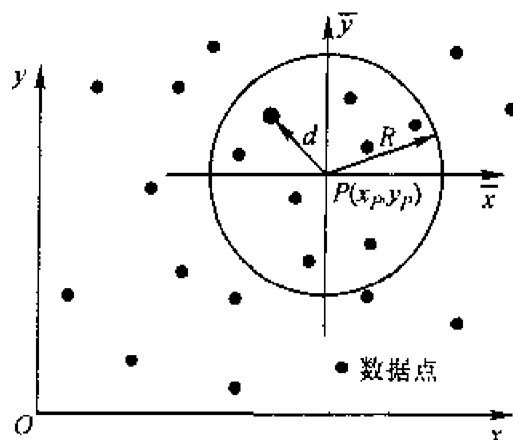


图 3-29 移动拟合方法

对采集的离散高程点，在插值运算之前，须进行认真的检查，检查其点位与记录的高程值是否匹配一致，点位是否有偏离现象，特征点高程是否有遗漏，有否重复记录的情况等等。检查工作可以在图形终端上进行，与原图进行核对，也可以通过打表输出检查，直至获得净化的取样数据文件，便可以转入插值运算。

2. 数据内插

由于取样的数据点呈离散分布形式，或者数据点虽按格网排列，但格网的密度不能满足使用的要求，这样就需要以数据点为基础进行插值运算。插值运算的手段是选择一个合理的数学模型，利用已知点上的信息求出函数的待定系数。由于地表面形态千变万化，既无规律性又无重复性，取用一个低次多项式来拟合整个地表面形态是不切合实际的。若采用高次多项式模拟地表面，又会出现函数的不稳定。因此，通常采用局部函数内插和逐点内插两种方式。

(1) 移动拟合法。移动拟合法是典型的逐点内插法。对每一个待定点取用一个多项式曲面拟合该点附近的地表面，此时取待定点作平面坐标的原点，并用待定点为圆心，以 R 为半径的圆内诸数据点来定义函数的待定系数，如图 3-29 所示。

设取二次多项式来拟合，则待求点的高程可写成一般式为

$$z_p = Ax^2 + Bxy + Cy^2 + Dx + Ey + F$$

将坐标原点平移到待定点处，即

$$\bar{x} = x - x_p \qquad \bar{y} = y - y_p$$

代入得到移动拟合法二次多项式插值公式

$$z_p = a\bar{x}^2 + b\bar{x}\bar{y} + c\bar{y}^2 + d\bar{x} + e\bar{y} + f$$

式中：$\bar{x}$、$\bar{y}$为数据点坐标经坐标系原点平移至待求点坐标 x_p、y_p 后的坐标值。

式中有 6 个待求参数，最少取用 6 个数据点的信息($\bar{x}$、$\bar{y}$、$\bar{z}$)代入求得。对于待求点$\bar{x}_p = \bar{y}_p = 0$；式中 F 值即为待求点的内插高程。当取用的数据点多于 6

个点时，则以数据点高程 z 作为观测值，列出误差方程式

$$Q = A\bar{x}^2 + B\bar{x}\bar{y} + C\bar{y}^2 + D\bar{x} + E\bar{y} + F - Z$$

并以数据点到待求点的距离给予适当的权，按最小二乘法平差原理求解。权的值应与距离成反比，间距愈近，对待求点测定值的影响应愈大。如取

$$W = \frac{1}{d^2} \text{或} \left(\frac{R-d}{d}\right)^2$$

式中：d 为待定点到数据点间的水平距离；

R 为定义函数待定参数时所求的圆半径。

该方法称为按距离加权最小二乘内插算法。

(2) 局部函数法。通常以格网小块为加密区，采用低次项函数拟合地表面。

① 线性内插。数据点位于地形特征点时，地面模型则以三角网形式建立。此时认为分块插值区的地表面为一平面，按直线比例内插待定点的高程，使用最靠近待定点的 3 个数据点来定义函数。设待定点高程为 z_p，则线性内插函数为

$$z_p = a_0 + a_1 x + a_2 y$$

将 3 个数据点的坐标值代入，联立求解出系数 a_0、a_1 和 a_2，待定点在给予平面坐标 x_p、y_p 之后，即可求出内插高程 z_p。

② 双线性多项式内插。认为分块插值区内待定点的高程在轴 x(和 y)平行的方向上与坐标 y(和 x)成直线比例关系。设待定点的高程为 z_p，则按双线性多项式内插的函数式为

$$z_p = a_0 + a_1 x + a_2 y + a_3 xy$$

此时取用最靠近的 4 个数据点定义函数的待定参数。

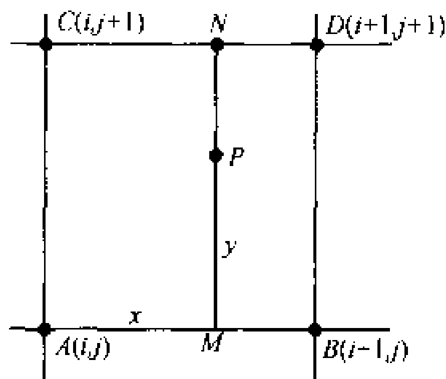


图 3－30　双线性内插

如果数据是按正方形格网节点布置，如图 3－30 所示，这时可先用点 $A(i,j)$ 和 $B(i+1,j)$ 及 $C(i,j+1)$ 和 $D(i+1,j+1)$ 两对点的高程，线性内插出点 M 和 N 的高程（z_M 和 z_N），然后再由 z_M 和 z_N 直线内插待求点 P 的高程 z_p。内插公式为

$$z_p = z(i,j)\left(1-\frac{x}{L}\right)\left(1-\frac{y}{L}\right) + z(i+1,j)$$

$$\left(1-\frac{y}{L}\right)\left(\frac{x}{L}\right) + z(i,j+1)\left(1-\frac{x}{L}\right)\left(\frac{y}{L}\right) + z(i+1,j+1)\left(\frac{x}{L}\right)\left(\frac{y}{L}\right)$$

式中：x、y 为以点 $A(i,j)$ 为坐标原点的待定点坐标；

L 为正方形格网的间距长度，当取格网间距 L 为单位长度 1 时，格网分块内加密的待定点也是正方形格网，则

$$z_p=(1-x)(1-y)z(i,j)+x(1-y)z(i+1,j)$$
$$+(1-x)yz(i,j+1)+xyz(i+1,j+1)$$

其中待定点坐标仍以点 $A(i,j)$为坐标原点，而

$$x=\left(\frac{1}{n}\right)(i-1)\quad i=1,2,3,\cdots,n;$$

$$y=\left(\frac{1}{n}\right)(j-1)\quad j=1,2,3,\cdots,n;$$

$n\times n$ 为内插正方形格网的数目，而待定点坐标为

$$0\leqslant x\leqslant 1\qquad\qquad 0\leqslant y\leqslant 1$$

③ 双三次多项式(样条函数)。在分块插值区用双三次多项式即样条函数模拟地表面。待定点高程为

$$z_p=f(x,y)=a_1x^3y^3+a_2x^2y^3+a_3xy^3+a_4y^3$$
$$+a_5x^3y^2+a_6x^2y^2+a_7xy^2+a_8y^2+a_9x^3y$$
$$+a_{10}x^2y+a_{11}xy+a_{12}y+a_{13}x^3+a_{14}x^2+a_{15}x+a_{16}\quad(3-13)$$

设数据点按正方形格网排列，每一格网作为分块单元，取格网间隔为单位长度，并取左下角的数据点为坐标原点，待定点坐标 $0\leqslant x\leqslant 1, 0\leqslant y\leqslant 1$。由于分块单元上 4 个格网节点的信息$(x,y,z)$只能列出 4 个方程式，而函数的待定参数却共有 16 个，因此把各数据点处在 x 方向的斜率 $R=\partial z/\partial x$，y 方向的斜率 $S=\partial z/\partial y$ 和曲面的扭曲 $T=\partial^2 z/\partial x\cdot\partial y$ 参与来定义函数的待定值。这样每一个数据点就可列出 4 个方程，4 个数据点就能解求出 16 个待定参数 a_i。某数据点的斜率 R、S 和 T 可以借助数据点的 4 个相邻网格上数据点的高程来推算。如图 3-31 所示的数据点 A

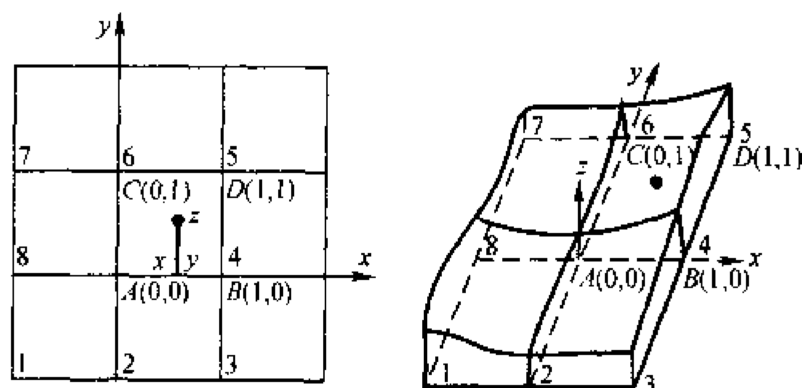


图 3-31 样条函数内插方法

(据崔炳光等，1984)

$$R_A = \partial z / \partial x = \frac{z_4 - z_8}{2}$$

$$S_A = \partial z / \partial y = \frac{z_6 - z_2}{2}$$

$$T_A = \frac{\partial}{\partial x}\left(\frac{\partial z}{\partial y}\right) = \frac{\partial}{\partial x}\left(\frac{z_6 - z_2}{2}\right)$$

而 $\frac{\partial z_6}{\partial x} = \frac{z_5 - z_7}{2}$； $\frac{\partial z_2}{\partial x} = \frac{z_3 - z_1}{2}$

得 $T_A = \frac{1}{2}\left(\frac{z_5 - z_7}{2} - \frac{z_3 - z_1}{2}\right) = \frac{1}{4}[(z_1 + z_5) - (z_3 + z_7)]$

因此4个数据点 A、B、C、D 列出 Z、R、S 和 T 的式子写成

$$\begin{bmatrix} Z_A \\ Z_B \\ Z_C \\ Z_D \\ R_A \\ R_B \\ R_C \\ R_D \\ S_A \\ S_B \\ S_C \\ S_D \\ T_A \\ T_B \\ T_C \\ T_D \end{bmatrix} = \begin{bmatrix} 0&0&0&0&0&0&0&0&0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&1&1&1&1 \\ 0&0&0&1&0&0&0&1&0&0&0&1&0&0&0&1 \\ 1&1&1&1&1&1&1&1&1&1&1&1&1&1&1&1 \\ 0&0&0&0&0&0&0&0&0&0&0&0&0&0&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&0&3&2&1&0 \\ 0&0&1&0&0&0&1&0&0&0&1&0&0&0&1&0 \\ 3&2&1&0&3&2&1&0&3&2&1&0&3&2&1&0 \\ 0&0&0&0&0&0&0&0&0&0&0&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&1&1&1&1&0&0&0&0 \\ 0&0&0&3&0&0&0&2&0&0&0&1&0&0&0&0 \\ 3&3&3&3&2&2&2&2&1&1&1&1&0&0&0&0 \\ 0&0&0&0&0&0&0&0&0&0&1&0&0&0&0&0 \\ 0&0&0&0&0&0&0&0&3&2&1&0&0&0&0&0 \\ 0&0&3&0&0&0&2&0&0&0&1&0&0&0&0&0 \\ 9&6&3&0&6&4&2&0&3&2&1&0&0&0&0&0 \end{bmatrix} \times \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ a_5 \\ a_6 \\ a_7 \\ a_8 \\ a_9 \\ a_{10} \\ a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \\ a_{15} \\ a_{16} \end{bmatrix}$$

写成矩阵形式：$Z = K \cdot A$，求出待定参数 $A = K^{-1} Z$，而

$K^{-1} =$

$$\begin{bmatrix}
4 & -4 & -4 & 4 & 2 & 2 & -2 & -2 & 2 & -2 & 2 & -2 & 1 & 1 & 1 & 1 \\
-6 & 6 & 6 & -6 & -4 & -2 & 4 & 2 & -3 & 3 & -3 & 3 & -2 & -1 & -2 & -1 \\
0 & 0 & 0 & 0 & 2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
2 & 0 & -2 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
-6 & 6 & 6 & -6 & -3 & -3 & 3 & 3 & -4 & 4 & -2 & 2 & -2 & -2 & -1 & -1 \\
9 & -9 & -9 & 9 & 6 & 3 & -6 & -3 & 6 & -6 & 3 & -3 & 4 & 2 & 2 & 1 \\
0 & 0 & 0 & 0 & -3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & 1 & 0 \\
-3 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & -2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & - & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -3 & 3 & 0 & 0 & -2 & -1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
2 & -2 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-3 & 3 & 0 & 0 & -2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}$$

如在分块格网 ABCD 内内插 $n \times n$ 个正方形待定点，则式(3－13)中的

$$x = \frac{1}{n}(i-1),\ i = 1,2,3,\cdots,n$$

$$y = \frac{1}{n}(j-1),\ j = 1,2,3,\cdots,n$$

双三次多项式样条函数是正方形格网布点时内插加密常用方法之一。

3．数据记录

数据记录是将内插的结果(格网或三角网数据)，按一定的数据结构形式，存储于磁性介质内，以供系统应用。为保证存储数据的正确有效，须对建立的数字高程模型进行精度分析。

精度分析的方法是，在研究区内随机选取 n 个数据点，设这些点的图上高程值为 z_i，其对应的内插值为 z_i'，它们之差用 $\Delta z_i = z_i - z_i'$ 表示，其算术平均值为

$$z = \frac{\sum_{i=1}^{n} \Delta z_i}{n}$$

则标准偏差　$D = \sqrt{\sum_{i=1}^{n} (\Delta z_i - z)/n}$。根据高程数据精度的评价标准，其允许误差($\Delta z$)应小于原始数据比例尺等高距的 1/3。若符合该条件，即认为所建立的数字高程模型满足精度要求，则可以作为数字地形模型分析的依据(详见第 5

章§1),或者分别按照格网或三角网(图3－32)的数据组织方法建立数字地形模型数据库。

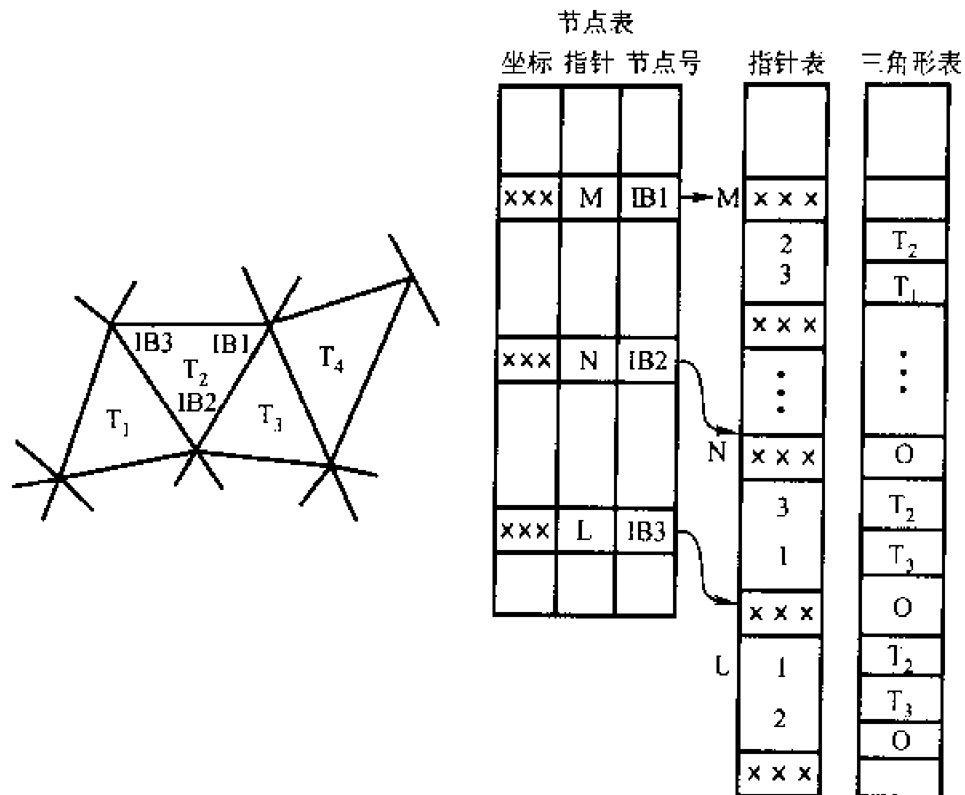


图3－32　数字高程模型的存储结构

5.2　区域的内插

区域的内插是研究根据一组分区的已知数据来推求同一地区另一组分区未知数据的内插方法。如图3－33所示,设一组已知数据的分区称为源区,需要内插的另一组分区称为目标区,则根据源

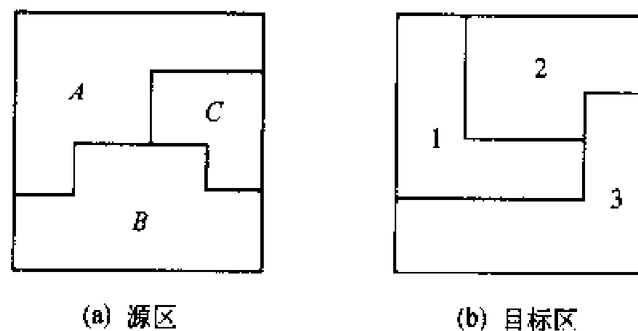


图3－33　区域内插示意图

区的数据来推求目标区的数据,可以采用以下两种方法。

1. 叠置法

叠置法是将目标区叠置在源区上，首先确定两者面积的交集 a_{ts}，然后利用下式算出目标区各个分区 t 的内插值 v_t。

$$v_t = \sum_s U_s a_{ts} / \sigma_s$$

式中：t 为目标区各个分区的序号；

s 为源区各个分区的序号；

U_s 为分区 s 的已知统计数据；

a_{ts} 为 t 区与 s 区相交的面积；

σ_s 为 s 区的面积。

例如，图 3-33 源区的已知数据如表 3-1 所示，目标区与源区叠置后得到的面积交集如表 3-2 所示，则 1 区的插值为

$$v_1 = 35 \times 3/7 + 30 \times 2/6 = 25$$

表 3-1　源区的人口和面积

	人　口	面　积
A	35	7
B	30	6
C	10	3

表 3-2　目标区与源区面积的交集

	A	B	C
1	3	2	0
2	4	0	1
3	0	4	2

以上，U_s 代表源区分区单元 s 的绝对统计指标，例如总人口、总收入等。如果 U_s 表示密度数据，例如人口密度等，则须先将密度值乘以 σ_s，转换为绝对统计数据，然后再将计算结果除以目标分区单元 t 的面积 r_t，其计算公式为

$$v_t = \left(\sum_s U_s \sigma_s a_{ts} / \sigma_s \right) / r_t = \sum_t U_s a_{ts} / r_t$$

同理，如果区域统计数据为比值，例如男性占总人口的百分比，则内插公式为

$$v_t = \frac{\sum_t U_{s1} a_{ts} / \sigma_s}{\sum_t U_{s2} a_{ts} / \sigma_s}$$

式中：$U_s = U_{s1} / U_{s2}$，U_{s1} 和 U_{s2} 为某一现象局部与整体的绝对统计数据或密度值。

2. 比重法

比重法是根据平滑密度函数的原理，将源区的统计数据从同质性改变为非同质性，而非同质性代表着一般社会经济现象的普遍特点。设以图 3-33 的已知数据为例，说明比重法的内插算法过程。

(1) 在源区上叠置一张格网，格网尺寸的大小应保证具有足够的内插精度。

(2) 将源区各个分区的平均人口数赋予相应分区的各个格网点，如图 3-

34(a)所示。

(3) 按公式

$$z_{i,j}=(z_{i-1,j}+z_{i+1,j}+z_{i,j+1}+z_{i,j+1})/4$$

计算相邻的四个格网点的平均值,得到图3-34(b)。

(4) 将各个分区的格网点值相加,设为 U_s',计算其系数 $p=U_s/U_s'$,并将各个格网点值乘以 p,得到调整后的各个分区的格网点值,如图3-34(c)所示。

(5) 依此过程继续下去,直到 U_s' 与 U_s 的值很相近,或者相应分区的格网点值比较一致时(图3-34d),便可计算目标区的内插值。例如,区1的内插值

$$v_1=5.5+5.3+5.5+5.2+4.5=26$$

同理,可求得目标区其他任意分区的内插值。

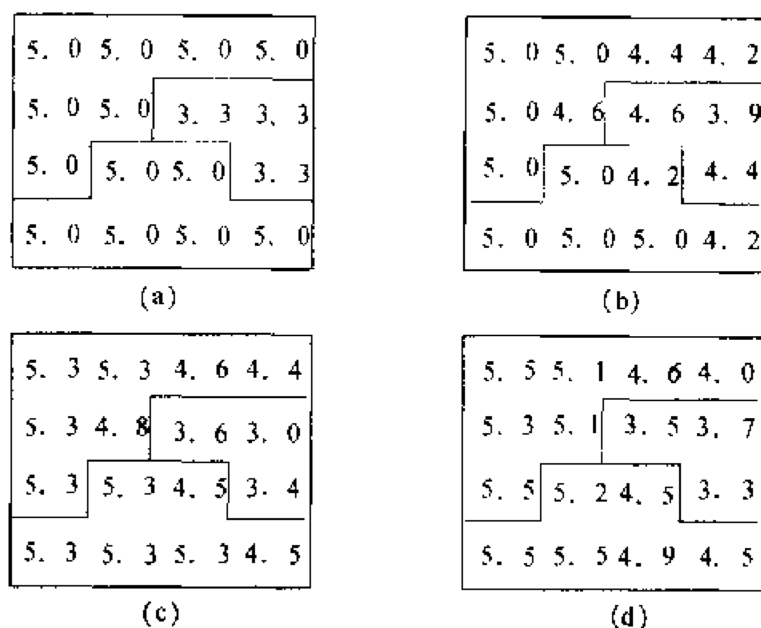


图3-34　比重法区域内插过程

§6　图幅数据边沿匹配处理

基于通常要将图幅的数字化数据连接在一起,组成连续统一的图幅数据,以便加入大型数据库或输出拼接后的图形(图3-35),这时,在相邻图幅的边沿部分,由于原图本身的数字化的误差,使得同一实体的线段或弧的坐标数据不能相互衔接,而且由于图幅的数字化数据,不仅坐标系统不一致,编码方法也不统一。因此,要将这些分幅数据联系在一起,组成统一的数据文件,需要进行图幅数据

边沿匹配的处理。边沿匹配处理的任务包括以下几个方面。

6.1　识别和检索相邻图幅的数据

每次对图幅数据的边沿匹配总是在相邻两幅图之间进行的，要将相邻两幅图的数据集中在一起，首先要识别它们，这可以通过图幅编号与分幅数字化的数据联系起来的方法来解决。如图3-36所示，为了便于计算机处理，各图幅常给以编号，其中十位数指示图幅编号的横向顺序，个位数指示纵向顺序，并记录图幅的长宽标准尺寸。因此，当进行横向图幅拼接时，总是将十位数编号相同的图幅数据收集在一起；当进行纵向图幅拼接时，是将个位数编号相同的图幅数据收集在一起。其次，图幅数据的边沿匹配处理主要是针对跨越相邻图幅的线段或弧而言的，为了减少数据容量，提高处理速度，一般只提取图幅边界2 cm范围内（如图3-36编号为22的斜线部分）的数据作为处理和匹配的目标，而且必须在各图幅的坐标数据按照本章§1提供的方法才能有效地作为匹配处理的工作文件。

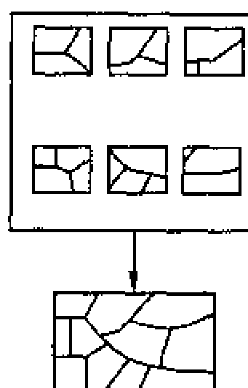


图3-35　图幅数据边沿匹配示意图

31	32	33
21	22	23
11	12	13

图3-36　图幅编号及图幅边沿数据的提取

6.2　相邻图幅边界点坐标数据的匹配

相邻图幅边界点坐标数据的匹配采用追踪拼接法。追踪拼接有四种情况（图3-37），只要符合下述条件，两条线段或弧即可匹配衔接：①相邻图幅边界两条线段或弧的左右码各自相同或相反；②相邻图幅同名边界点坐标在某一许可定值的范围内（例如≤±0.5 mm）。匹配衔接时是以一条弧或线段链作为处理的单元，因此，当边界点位于两个结点之间时，须分别取出相关的两个结点，然后按照结点之间线段链方向一致性的原则进行数据的记录和存储。例如，图3-39就是根据图3-38的原始数据，按照跟踪拼接法处理后输出的

图形。

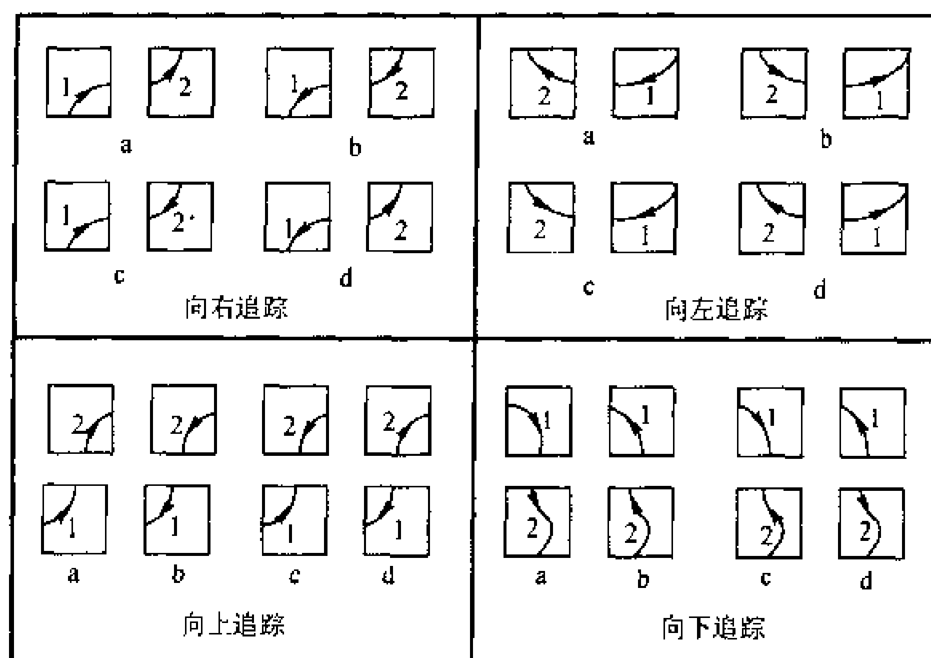


图 3-37 追踪拼接法
(据胡友元等,1988)

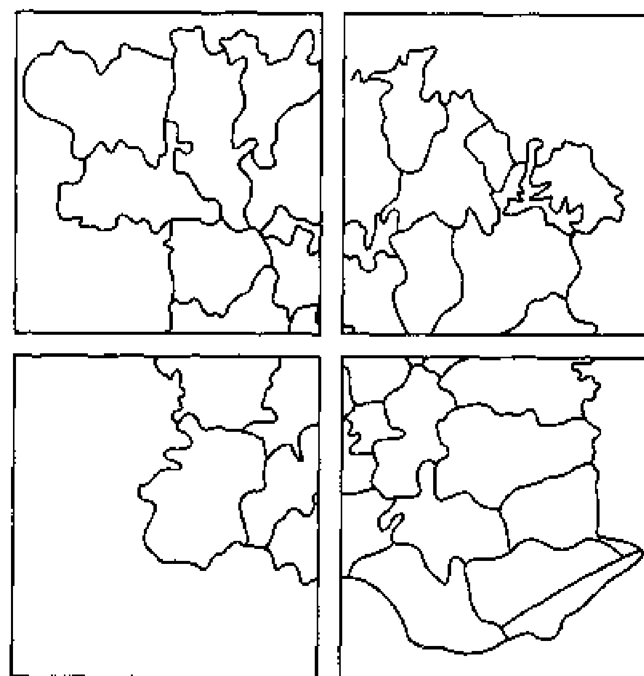


图 3-38 图幅边沿匹配的资料原图
(据胡友元等,1988)

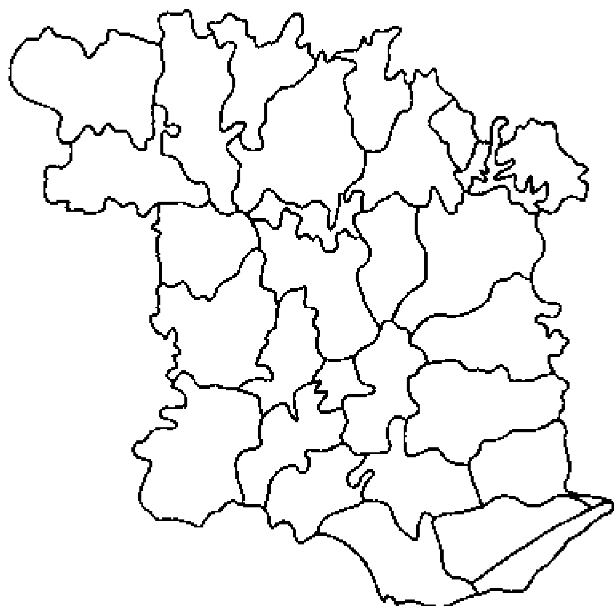


图3-39　图幅边沿匹配后的境界线
（据胡友元等，1988）

6.3　相同属性多边形公共界线的删除

当图幅数据通过边沿匹配处理，组成较大区域的连续图幅的数据以后，常常伴随着比例尺的缩小和属性类型的合并，需要对具有相同属性的多边形公共边界线予以删除。删除的方法参阅本章§4之4.2。

思考题

1. 讨论在空间数据坐标变换中，选择控制点数量及分布的判断标准。

2. 试设计一算法，使图形数字化作业员两次放置地图原图在数字化仪平台上，数字化操作后得到的坐标数据仍处于同一坐标系。

3. 在下列环境中你将使用什么类型的投影？为什么？

① 一位飞行员在从New York到London的北大西洋上空飞行；

② 一艘潜水艇在北极的冰山下航行；

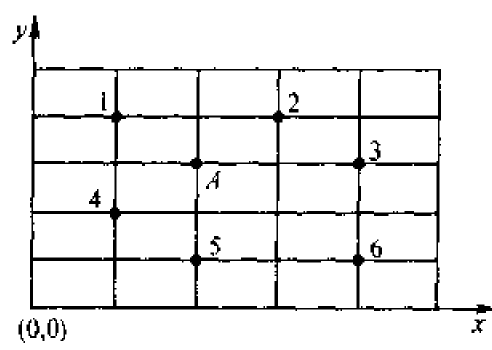
③ 一位农业科学家在汇编非洲的农作物产量数据；

④ 一名工程师在规划横跨中国全境的无线电发报机的分布位置。

4. 比较空间内插的移动拟合法、局部函数和按距离加权法等。如图3-40，试用一次多项式拟合法和按距离加权法分别求取A点的高程，并比较其结果。

5. 将由空间插值法绘制的等值线图与有经验的制图人员绘制的等值线图作比较，检验由空间插值法绘制的等值线图的效果和可能出现的问题。

6. 除本章§2所述的空间数据结构转换方法外，你还了解其他哪些方法？比较它们的优缺点和使用的场合。



点号	x/cm	y/cm	z/m
1	1	4	40
2	3	4	24
3	4	3	30
4	1	2	24
5	2	1	32
6	4	1	25

图 3-40

第4章 地理信息系统空间数据库

§1 空间数据库概述

数据库是一个信息系统的基本且重要的组成部分。同样，在一个空间项目的工作流程中，地理信息系统负责管理工作流程所需的空间数据，辅助用户用这些数据进行空间分析、综合，进而进行所需的决策。所以，地理信息系统是该项目工作过程的支柱。空间数据库是地理信息系统中空间数据的存储场所。在一个项目的工作过程中，空间数据库发挥着核心的作用。这集中表现在：用户在决策过程中，通过访问空间数据库获得空间数据，在决策过程完成后再将决策结果存储到空间数据库中。空间数据库的布局和存取能力对地理信息系统功能的实现和工作的效率影响极大。如果在组织的所有工作地点都能很容易地存取各种数据，则能使地理信息系统快速响应组织内决策人员的要求；反之，就往往会妨碍地理信息系统的快速响应。如果获取空间数据很困难，就不可能进行及时的决策，或者只能根据不完全的空间数据进行决策，其结果都可能导致地理信息系统不能得出正确的决策结果。可见空间数据库在地理信息系统中的重要性是不言而喻的。

1.1 空间数据库的概念

通常，数据库是数据库系统的简称。一个完整的数据库系统应该包括数据库、数据库管理系统（Database Management System, DBMS）和数据库应用系统三个组成部分。其中，数据库是按照一定的结构组织在一起的相关数据的集合；数据库管理系统是提供数据库建立、使用和管理工具的软件系统；而数据库应用系统则是为了满足特定的用户数据处理需求而建立起来的，具有数据库访问功能的应用软件，它提供给用户一个访问和操作特定数据库的用户界面。

同理，空间数据库系统也是由上述三个部分所组成。其中空间数据库指的是地理信息系统在计算机物理存储介质上存储的与应用相关的地理空间数据的总和，一般是以一系列特定结构的文件的形式组织在存储介质之上的。空间数据库管理系统则是指能够对物理介质上存储的地理空间数据进行语义和逻辑上

的定义，提供必需的空间数据查询检索和存取功能，以及能够对空间数据进行有效的维护和更新的一套软件系统。空间数据库管理系统的实现是建立在常规的数据库管理系统之上的。它除了需要完成常规数据库管理系统所必备的功能之外，还需要提供特定的针对空间数据的管理功能。常常有两种空间数据库管理系统的实现方法，一是直接对常规数据库管理系统进行功能扩展，加入一定数量的空间数据存储与管理功能。运用这一种方法比较有代表性的是 Oracle 等系统。另一种方法是在常规数据库管理系统之上添加一层空间数据库引擎，以获得常规数据库管理系统功能之外的空间数据存储和管理的能力。代表性的系统是 ESRI 的 SDE(Spatial Database Engine)等。由地理信息系统的空间分析模型和应用模型所组成的软件可以看作是空间数据库系统的数据库应用系统，通过它不但可以全面地管理空间数据，还可以运用空间数据进行分析与决策。

由此可见，空间数据库系统在整个地理信息系统中占有极其重要的地位，是地理信息系统发挥作用的关键。空间数据库设计的成败，直接影响到地理信息系统开发与应用水平及成效。

1.2 空间数据库的设计

空间数据库的设计问题，其实质是将地理空间客体以一定的组织形式在数据库系统中加以表达的过程，也就是地理信息系统中空间客体数据的模型化问题。

1. 空间数据库设计过程

地理信息系统是人类认识客观世界、改造客观世界的有力工具。地理信息系统的开发和应用需要经历一个由现实世界到概念世界，再到计算机信息世界的转化过程。如图 4-1 所示。概念世界的建立是通过对错综复杂的现实世界的认识与抽象，即对各种不同专业领域的研究和系统分析，最终形成地理信息系统的空间数据库系统和应用系统所需的概念化模型。进一步的逻辑模型设计，其任务就是把概念模型结构转换为计算机数据库系统所能够支持的数据模型。逻辑模型设计时最好应选择对某个概念模型结构支持得最好的数据模型，然后再选定能支持这种数据模型，且最合适的数据库管理系统。最后的存储模型则是指概念模型反映到计算机物理存储介质中的数据组织形式。

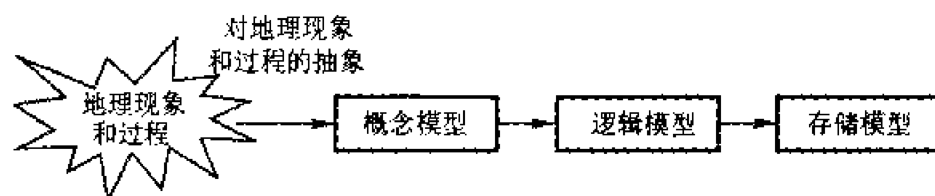


图 4-1 地理信息系统空间数据库模型的建立过程

地理信息系统的概念模型，是人们从计算机环境的角度出发和思考，对现实世界中各种地理现象、它们彼此的联系及其发展过程的认识及抽象的产物。具体地说，主要包括对地理现象和过程等客体的特征描述、关系分析和过程模拟等内容。这些内容在地理信息系统的软件工具、数据库系统和应用系统研究中往往被抽象、概括为数据结构的定义、数据模型的建立及专业应用模型的构建等主要理论与技术问题。它们共同构成地理信息系统基础研究的主要内容。

地理信息系统的空间数据结构是对地理空间客体所具有的特性的一些最基本的描述。地理空间是一个三维的空间，其空间特性表现为四个最基本的客体类型，即点、线、面和体等。这些客体类型的关系是十分复杂的。一方面，线可以视为由点组成，面可由作为边界的线所包围而形成，体又可以由面所包围而形成。可见四类空间客体之间存在着内在的联系，只是在构成上属于不同的层次。另一方面，随着观察这些客体的坐标系统的维数、视角及比例尺的变化，客体之间的关系和内容可能按照一定的规律相互转化。例如，由三维坐标系统变为二维坐标系统后，比如通过地图投影，空间体可变成面，面可以部分地变成线，线可以部分地变成点。视角变化之后，也将使某些客体发生变化。坐标系统的比例尺缩小时，部分的体、面、线客体均可能变为点客体。由此可见，空间点、线、面和体等客体及它们之间结构上的关系是地理信息系统空间数据结构的基础。

同时，所有地理现象和地理过程中的各种空间客体并非孤立存在，而是具有各种复杂的联系。这些联系可以从空间客体的空间、时间和属性三个方面加以考察。

(1) 客体间的空间联系大体上可以分解为空间位置、空间分布、空间形态、空间关系、空间相关、空间统计、空间趋势、空间对比和空间运动等等联系形式。其中，空间位置描述的是空间客体个体的定位信息；空间分布是描述空间客体的群体定位信息，且通常能够从空间概率、空间结构、空间聚类、离散度和空间延展等方面予以描述；空间形态反映空间客体的形状和结构；空间关系是基于位置和形态的实体关系；空间相关是空间客体基于属性数据上的关系；空间统计描述空间客体的数量、质量信息，又称为空间计量；空间趋势反映客体空间分布的总体变化规律；空间对比可以体现在数量、质量和形态三个方面；空间运动则反映空间客体随时间的迁移或变化。以上种种空间信息基本上反映了空间分析所能揭示的信息内涵，彼此互有区别又有联系。

(2) 客体之间的时间联系一般可以通过客体变化过程来反映。有些客体数据的变化周期很长，如地质地貌等数据随时间的变化。而有些空间数据则变化很快，需要及时更新，如土地利用数据等。客体时间信息的表达和处理构成了空间时态地理信息系统及其数据库的基本内容。

(3) 客体间的属性联系主要体现为属性多级分类体系中的从属关系、聚类

关系和相关关系。从属关系主要反映各客体之间的上下级或包含关系；聚类关系是反映客体之间的相似程度及并行关系；相关关系则反映不同类客体之间的某种直接或间接的并发或共生关系。属性联系可以通过地理信息系统属性数据库的设计加以实现。

2. 空间数据库的数据模型设计

对于上述地理空间客体及其联系的数学描述，可以用数据模型这个概念进行概括。建立空间数据库系统数据模型的目的，是揭示空间客体的本质特性，并对其进行抽象化，使之转化为计算机能够接受和处理的数据形式。在地理信息系统研究中，空间数据模型就是对空间客体进行描述和表达的数学手段，使之能反映客体的某些结构特性和行为功能。按数据模型组织的空间数据使得数据库管理系统能够对空间数据进行统一的管理，帮助用户查询、检索、增删和修改数据，保障空间数据的独立性、完整性和安全性，以利于改善对空间数据资源的使用和管理。空间数据模型是衡量地理信息系统功能强弱与优劣的主要因素之一。数据组织的好坏直接影响到空间数据库中数据查询、检索的方式、速度和效率。从这一意义上看，空间数据库的设计最终可以归结为空间数据模型的设计。

数据库系统中通常采用的数据模型主要有层次模型、网状模型和关系模型，以及语义模型、面向对象的数据模型等。这些数据模型都可以用于空间数据库的设计。

3. 空间数据库设计的原则、步骤和技术方法

随着地理信息系统空间数据库技术的发展，空间数据库所能表达的空间对象日益复杂，数据库和用户功能日益集成化，从而对空间数据库的设计过程提出了更高的要求。许多早期的空间数据库设计过程着重强调的是数据库的物理实现，注重于数据记录的存储和存取方法。设计人员往往只需要考虑系统各个单项独立功能的实现，从而也只考虑少数几个数据库文件的组织，然后选择适当的索引技术，以满足实现这个功能时的性能要求。而现在，对空间数据库的设计已提出许多准则，其中包括：① 尽量减少空间数据存储的冗余量；② 提供稳定的空间数据结构，在用户的需要改变时，该数据结构能迅速作相应的变化；③ 满足用户对空间数据及时访问的需求，并能高效地提供用户所需的空间数据查询结果；④ 在数据元素间维持复杂的联系，以反映空间数据的复杂性；⑤ 支持多种多样的决策需要，具有较强的应用适应性。

地理信息系统数据库设计往往是一件相当复杂的任务，为有效地完成这一任务特别需要一些合适的技术，同时还要求将这些设计技术正确组织起来，构成一个有序的设计过程。设计技术和设计过程是有区别的。设计技术是指数据库设计者所使用的设计工具，其中包括各种算法、文本化方法、用户组织的图形表示法、各种转化规则、数据库定义的方法及编程技术；而设计过程则确定了这些

技术的使用顺序。例如,在一个规范的设计过程中,可能要求设计人员首先用图形表示用户数据,再使用转换规则生成数据库结构,下一步再用某些确定的算法优化这一结构,这些工作完成后,就可进行数据库的定义工作和程序开发工作。

一般说来,数据库设计技术分为下列两类:① 数据分析技术,数据分析技术是用于分析用户数据的语义的技术手段;② 技术设计技术,技术设计技术用于将数据分析结果转化为数据库的技术实现。

上述两类技术所处理的是两类不同的问题。第一类问题考虑的是正确的结构数据,这些问题通过使用诸如消除数据冗余技术、保证数据库稳定性技术、结构数据技术来解决,其目的是使用户易于存取数据,从而满足用户对数据的各种需求。第二类问题是保证所实现的数据库能有效地使用数据资源,解决这个问题要用到一些技术设计技术,例如选择合适的存储结构以及采用有效的存取方法等等。

数据库设计的内容包括了数据模型的三个方面:即数据结构、数据操作和完整性约束。具体区分为:① 静态特性设计,又称结构特性设计。也就是根据给定的应用环境,设计数据库的数据模型(即数据结构)或数据库模式。它包括概念结构设计和逻辑结构设计两个方面。② 动态特性设计,又称数据库的行为特性设计。设计数据库的查询、静态事务处理和报表处理等应用程序;③ 物理设计。根据动态特性,即应用处理要求,在选定的数据库管理系统环境之下,把静态特性设计中得到的数据库模式加以物理实现,即设计数据库的存储模式和存取方法。

数据库设计的整个过程包括以下几个典型步骤,在设计的不同的阶段要考虑不同的问题,每类问题有其不同的自然论域。在每个设计阶段必须选择适当的论述方法及与其相应的设计技术。这种方法强调的是,首先将确定用户需求与完成技术设计相互独立开来,而对其中每一个大的设计阶段再划分为若干更细的设计步骤,如图 4-2 所示:

(1) 需求分析。即用系统的观点分析与某一特定的数据库应用有关的数据集合。

(2) 概念设计。把用户的需求加以解释,并用概念模型表达出来。概念模型是现实世界到信息世界的抽象,具有独立于具体的数据库实现的优点,因此是用户和数据库设计人员之间进行交流的语言。数据库需求分析和概念设计阶段需要建立数据库的数据模型,可采用的建模技术方法主要有三类:一是面向记录的传统数据模型,包括层次模型、网状模型和关系模型;二是注重描述数据及其之间语义关系的语义数据模型,如实体-联系模型等;三是面向对象的数据模型,它是在前两类数据模型的基础上发展起来的面向对象的数据库建模技术。本章将依次论述这些模型在空间数据库设计中的应用,并将数据库实现模型中

的一些存储方法及查询技术一并加以阐述。

(3) 逻辑设计。数据库逻辑设计的任务是:把信息世界中的概念模型利用数据库管理系统所提供的工具映射为计算机世界中为数据库管理系统所支持的数据模型,并用数据描述语言表达出来。逻辑设计又称为数据模型映射。所以,逻辑设计是根据概念模型和数据库管理系统来选择的。例如将上述概念设计所获得的实体-联系模型转换成关系数据库模型。

(4) 物理设计。数据库的物理设计指数据库存储结构和存储路径的设计,即将数据库的逻辑模型在实际的物理存储设备上加以实现,从而建立一个具有较好性能的物理数据库。该过程依赖于给定的计算机系统。在这一阶段,设计人员需要考虑数据库的存储问题:即所有数据在硬件设备上的存储方式,管理和存取数据的软件系统,数据库存储结构以保证用户以其所熟悉的方式存取数据,以及数据在各个位置的分布方式等。

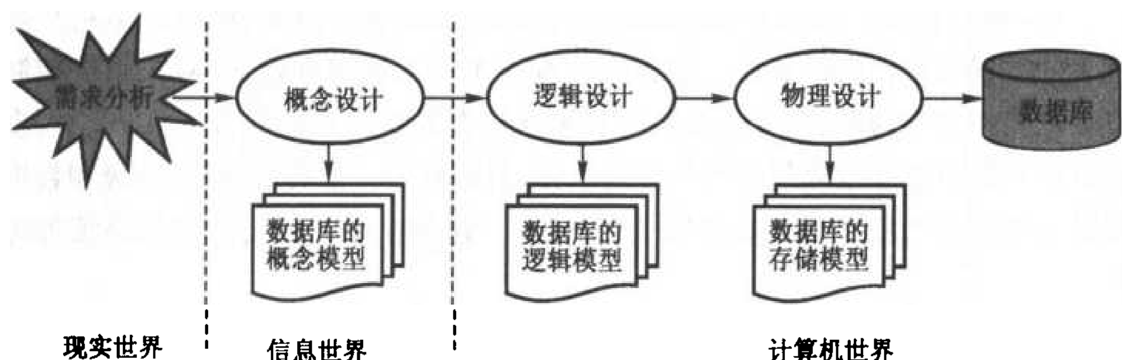


图4-2 数据库设计的步骤

1.3 空间数据库的实现和维护

1. 空间数据库的实现

根据空间数据库逻辑设计和物理设计的结果,就可以在计算机上创建起实际的空间数据库结构,装入空间数据,并测试和运行,这个过程就是空间数据库的实现过程,它包括:① 建立实际的空间数据库结构;② 装入试验性的空间数据对应用程序进行测试,以确认其功能和性能是否满足设计要求,并检查对数据库存储空间的占有情况;③ 装入实际的空间数据,即数据库的加载,建立起实际运行的空间数据库。

2. 相关的其他设计

其他设计的工作包括加强空间数据库的安全性、完整性控制,以及保证一致性、可恢复性等,总之是以牺牲数据库运行效率为代价的。设计人员的任务就是要在实现代价和尽可能多的功能之间进行合理的平衡。这一设计过程包括:

(1) 空间数据库的再组织设计。对空间数据库的概念、逻辑和物理结构的

改变称为再组织，其中改变概念或逻辑结构又称再构造，改变物理结构称为再格式化。再组织通常是由于环境需求的变化或性能原因而引起的。一般数据库管理系统，特别是关系型数据库管理系统都提供数据库再组织的实用程序。

(2) 故障恢复方案设计。在空间数据库设计中考虑的故障恢复方案，一般是基于数据库管理系统提供的故障恢复手段，如果数据库管理系统已经提供了完善的软硬件故障恢复和存储介质的故障恢复手段，那么设计阶段的任务就简化为确定系统登录的物理参数，如缓冲区个数、大小，逻辑块的长度，物理设备等。否则就要制订人工备份方案。

(3) 安全性考虑。许多数据库管理系统都有描述各种对象(记录，数据项)的存取权限的成分。在设计时根据用户需求分析，规定相应的存取权限。子模式是实现安全性要求的一个重要手段。也可在应用程序中设置密码，对不同的使用者给予一定的密码，以密码控制使用级别。

(4) 事务控制。大多数数据库管理系统都支持事务概念，以保证多用户环境下的数据完整性和一致性。事务控制有人工和系统两种控制办法，系统控制以数据操作语句为单位，人工控制则以事务的开始和结束语句显示实现。大多数数据库管理系统也提供封锁粒度的选择，封锁粒度一般有库级、记录级和数据项级。粒度越大控制越简单，但并发性能差。这些在相关的设计中都要统筹考虑。

3. 空间数据库的运行与维护

空间数据库投入正式运行，标志着数据库设计和应用开发工作的结束和运行维护阶段的开始。本阶段的主要工作是：① 维护空间数据库的安全性和完整性：需要及时调整授权和密码，转储及恢复数据库；② 监测并改善数据库性能：分析评估存储空间和响应时间，必要时进行数据库的再组织；③ 增加新的功能：对现有功能按用户需要进行扩充；④ 修改错误：包括程序和数据。

§2　空间数据库概念模型设计——传统的数据模型

数据库系统把相关的数据集合以集成的方法加以组织，使得用户能有效地管理和处理数据。实际运用中，特别是在涉及空间数据的应用中，数据的性质和特征相当复杂，结构和表达方式也相应具有复杂的变化形式。所以，数据库中需要有一些形式化的方法来描述数据的逻辑结构和各种操作，于是产生了数据模型的概念。数据模型可看作是表达现实世界的规格化的说明。具体地说，数据库的数据结构、操作集合和完整性约束规则集合等组成了数据库的数据模型。

传统的数据模型主要指层次、网状和关系三种模型，它们是计算机中以文件系统组织的数据模型的继承和发展。在地理信息系统发展的初期，空间数据的

组织主要是以文件的方式存储。随着数据库理论和技术的成熟，孤立的文件系统逐步被取代，简单的文件操作也发展成为较复杂的数据库操作和相应的数据语言，这无疑是在类似空间数据这样的大规模数据管理方面取得的一个重大技术进步，促进了地理信息系统的发展。

2.1 层次数据模型

层次数据模型描述了各类客体及客体类之间的联系。然而，层次模型限制每一个客体类最多只能有一个双亲客体类，而一个双亲客体则可有多个子女客体类。双亲和子女客体类之间即形成了所谓的层次关系。在层次数据模型中，记录被表示为树型结构的一个节点。客体间组织起双亲－子女关系。在此以地理信息系统中矢量多边形数据的组织为例，如图4－3所示，一个多边形记录(polygons)关联着若干个组成它的弧段记录(arcs)，则这些弧段就是属于这个多边形节点的子女节点，多边形节点是双亲节点。同样，一个弧段可以有两个分别作为开始和结束的端点记录(nodes)作为其子女节点。由于层次模型具有只允许有一个双亲的限制，由此就可以得出一个重要的推论：那些拥有多于一个双亲的客体必然在数据库中要重复出现多次。比如弧段 arc 1 是两个多边形 polygon 1 和 polygon 2 所共有的一条边界弧段，因而弧段 arc 1 就必须在数据库中出现两次。基于这一种原因，对于像 M:N 这样多对多性质的客体间的联系利用层次数据模型描述就会出现一些问题，因为这时数据库中需要重复存储这些数据，造成冗余。

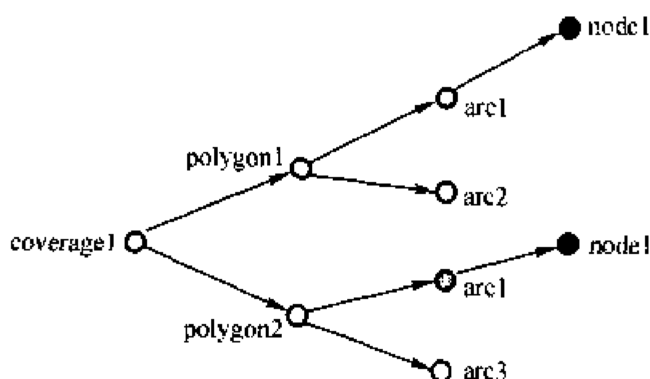


图4－3 多边形客体的层次

1. 层次数据结构

数据库的层次数据结构可以用如图4－4的数据结构图来形式化地表达。其中，客体类型用矩形框示出，双亲－子女联系则用从双亲到子女的有向线段表示。在图4－4所示的层次数据结构中，可能一个双亲客体类拥有多个子女客体类，如弧段 arcs 记录客体类的节点以 nodes 记录客体类和坐标 coordinates 记录客体类作为其子女节点，即客体 nodes 和 coordinates 是其双亲节点弧段 arcs 的同一

层次的子女节点。

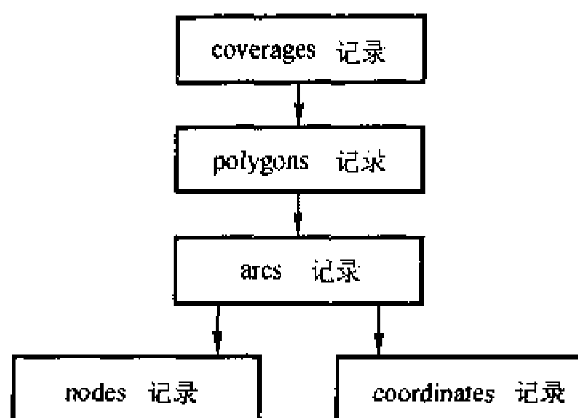


图4-4 多边形层次数据结构

2. 层次数据结构的数据存取

层次数据库中客体的存取有两种方法：树遍历法和通用选择法。

(1) 树遍历法。在这种存取方法中，客体是以某种树的遍历顺序检索的。通常采用前序遍历，即首先搜索双亲节点，然后搜索其子女节点的搜索顺序。在以前序方法进行树遍历检索时，如果某个双亲节点客体类拥有多个子女客体类，则先检索这一双亲节点的所有子女节点，然后再检索与这个双亲节点在同一级的其他双亲节点。又由于在一些层次数据库的实现中，同一双亲节点客体类的不同子女节点客体类的存储是有次序的，记录的检索就根据这种次序进行。比如 arcs 记录客体类中，nodes 记录客体类是左节点类，coordinates 记录客体类是右客体类。在每次检索操作中，nodes 记录总是先于 coordinates 记录被检索出来。

(2) 通用选择法。这种方法不依赖于客体在树结构中的顺序，而是根据所确定的选择条件，在结构中选择某特定的客体。例如：

DRAW coordinates WHERE arcs = ‘arc 1’；从数据库中提取弧段 arc 1 的坐标数据并显示。

DRAW polygons WHERE arcs = ‘arc 1’；显示包含弧段 arc 1 的所有多边形数据。

逻辑检索条件可能很复杂，可能包括关于客体本身的特性、不同客体特性的比较，以及客体集合比较布尔表达式等形式。

2.2 网状数据模型

网状数据模型以系结构为基础，系(set)的概念和数学中的集合(set)的概念不同，数学中的集合是由同一确定的客体类的客体所组成，而网状数据模型的系结构是由属于两个不同客体类(即首、属两个客体类)的客体所组成。每

个系中只包含一个属于首客体类的客体，以及包含若干个属于属客体类的客体。因此，通常用首耦合系(owner-coupled set)这个术语来区分网状模型中的系和数学中的集合。并且在网状模型文献中，常用系值(set occurence)这个名称代替系的使用。系值由首客体类的一个客体和属客体类的若干个(包含零个)客体组成。同一系的所有系值组合为系类。一个系类中的所有系值具有下列性质：① 系中的首客体同属于首客体类；② 系中的属客体同属于属客体类。

图 4-5 表示了 polygon-arc 系类的两个系值。其中，多边形 polygons 是一个客体类，与其关联的弧段 arcs 也是一个客体类。他们共同组成 polygon-arc 系类，多边形充当首客体类的角色，而弧段则作为多边形的属客体类。

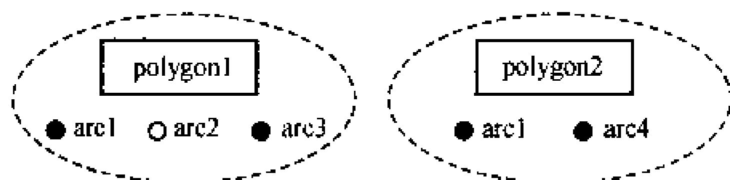


图 4-5 polygon-arc 系类的系值

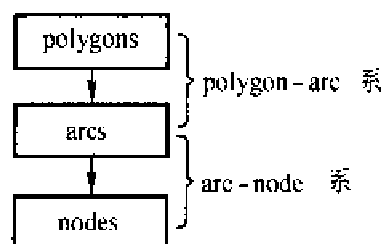


图 4-6 多边形的网状数据结构

1. 网状数据结构

地理信息系统中矢量多边形结构的网状数据模型可以用层次数据模型相同的数据结构图来表示。如图 4-6 所示，网状模型的首、属客体类都用矩形框来表示。首、属客体类的联系用一条从首客体类到属客体类的有向线段表示。在网状数据模型中并不限制客体类和系类的个数，也不限制客体类之间的从属关系，即某一个客体类可以作为其他任何一个系类的首客体，也可以作为其他任何一个系类的属客体。例如在图 4-6 中，客体类 arcs 在 polygon-arc 系类中是属客体类，而在 arc-node 系类中，却是首客体类。

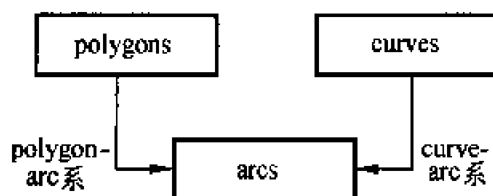


图 4-7 从属于多个首客体的客体

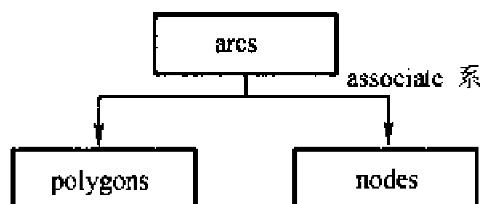


图 4-8 多属客体系类

另一种常见的系结构是某一客体出现在多个系类的系值中，即在这种情况下，某一客体可从属于多个首客体，这些首客体又分属于不同的系类。如图 4-7 所示，多边形客体类、线状客体类、以及弧段客体类是三种不同的客体类。其中，多边形客体类 polygons 与弧段客体类 arcs 组成 polygon-arc 系类，而线状客体

类 curves 与弧段客体类 arcs 组成 curve-arc 系类。arcs 客体类在两个不同的系类中充当属客体类。

网状数据模型的系结构还可以进一步扩展为更加复杂的结构，最为显著的扩展如下：

(1) 允许一个给定系类的各个属客体来自于不同的客体类，如图 4-8 所示，在弧段的关联系类 associate 中，弧段客体类 arcs 是首客体类，两个属客体类分别是多边形客体类 polygons 和弧段端点客体类 nodes，它们共同形成了地理信息系统的空间关联的拓扑结构。

(2) 允许一个系类的各个首客体可以从属于不同的客体类，如图 4-9 所示，地理信息系统空间索引系类 spatial-index 的首客体类可以是多边形客体类 polygons、弧段客体类 arcs 等，而它们具有共同的称为最小边界矩形的属客体类 MBRs。

(3) 上述两种情况的组合，即系值的首、属客体都可来自于不同的客体类。如图 4-10 所示，每个空间索引系 spatial-index 的首客体可以是一个多边形客体类 polygons，也可以是弧段客体类 arcs。而最小边界矩形客体类 MBRs 和线性四叉树空间索引代码客体类 quadtreecodes 都可以作为其属客体类。

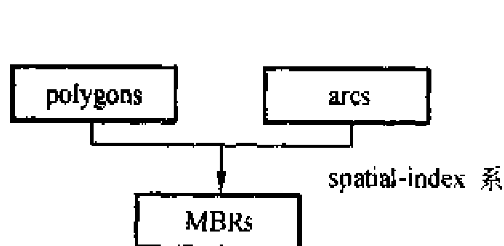


图 4-9　多个首客体的系类

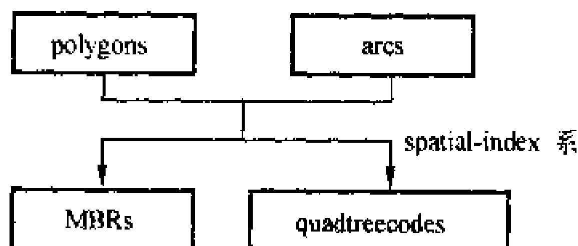


图 4-10　多首客体和多属客体的系类

2. 网状数据模型的实现

网状数据模型的实现以记录为基本存储结构，即每个客体都实现为一个记录结构。而系值的实现是将一个系值的各个记录链接起来。当以这种方式实现网状模型时，每一个客体类作为一个记录型，每个系类作为一个系型。图 4-5 至 4-10 的数据结构图既表示了网状模型，同时也表示了它的实现。实现时，图中的每个矩形框是记录型(record type)，每一条连接矩形框的线段称为系型(set type)。

一般来说，在系统实现时都对网状结构加以某些限制。大多数系统限定数据库设计人员选择这样的结构，即一个记录值只能作为同一个系型的某一个系值的属记录，这是由于系值是以链接方式实现的。如图 4-11 所示，系值的实现构成一个链，此链从该系值的首记录值开始，连接起所有属于此系值的属记录值。实现中每一个记录内部对其加入的每个系型相应只有一个存储指针的空

间，如某一记录属于同一系型的两个系值，如图 4-5 所示，则需要两个指针，分别指向两个系值中的后继记录值，但由于在记录中对一个系型只提供一个指针的存储空间，所以不能实现上述结构。

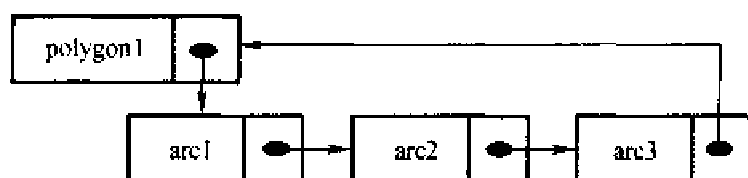


图 4-11 系型的链接实现

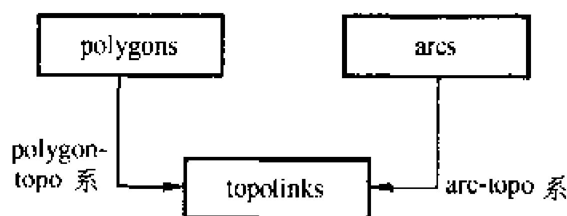


图 4-12 N:M 型联系的实现方法

要解决上述问题，并在网状数据库中实现 M:N 型的联系，就需要网状模型中用两个系型表示这种联系。如图 4-12 所示，在多边形客体类 polygons 和弧段客体类 arcs 之间建立一个组合的拓扑链接客体类 topolinks，可以用来存储组成多边形的弧段的排列顺序。由此形成了两个独立的系类，即多边形拓扑结构系 polygon-topo 和弧段拓扑结构系 arc-topo，使得 topolinks 记录型的每一个值只会在 polygon-topo 系型的一个系值中以及在 arc-topo 系型的一个系值中出现。M:N 型联系的链接实现方法如图 4-13 所示。

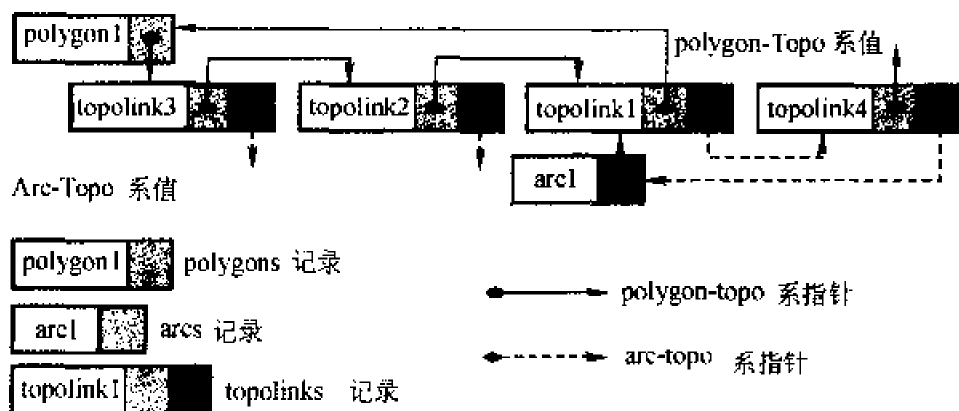


图 4-13 系型的链接实现

网状模型实现中的另一个限制是：不允许同一个记录型既作为一个系型的首记录，又作为该系型的属记录。例如多边形的拓扑包含关系中，多边形记录型就存在这种形式，如图 4-14(a)所示。实现这种结构可以通过引入一个新的纪录型 structures 来完成。对图 4-14(a)所示的拓扑包含关系的每一个组合都建

立一个 structures 记录值，并在其中存储多边形拓扑包含的层次数据，就可以实现这一联系的网状数据模型，如图 4－14(b)所示。而其链接的实现形式如图 4－14(c)所示，每个 polygons 和 structures 记录都需要具有两个系指针，实线表示的系指针代表 inclusion 系的联系，虚线表示的系指针代表 compose 系的联系。

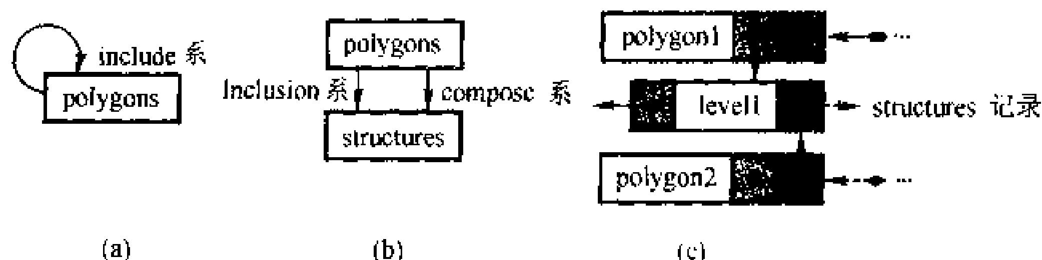


图 4－14 拓扑包含关系的网状模型

3．网状数据模型数据库的记录存取

在网状数据模型的数据库中经常用导航来描述存取中所用的方法，即存取可以从网状模型中的任何一个客体出发，沿着系指针到达其他的客体。如在图 4－13 所示的多边形和弧段网状模型中，要找到某个多边形所包含的所有弧段的信息，可按如下步骤进行：① 从该多边形的首记录值开始；② 沿着 polygon-topo 系指针查找到该多边形在 polygon-topo 系值中的所有属记录值；③ 对每个 polygon-topo 系值中的属记录值，沿着 arc-topo 系指针找出其在 arc-topo 系中的首记录值。

使用这种方法，可以从网状结构的任一点出发，利用系型的首、属记录联系找到任何相关的客体。

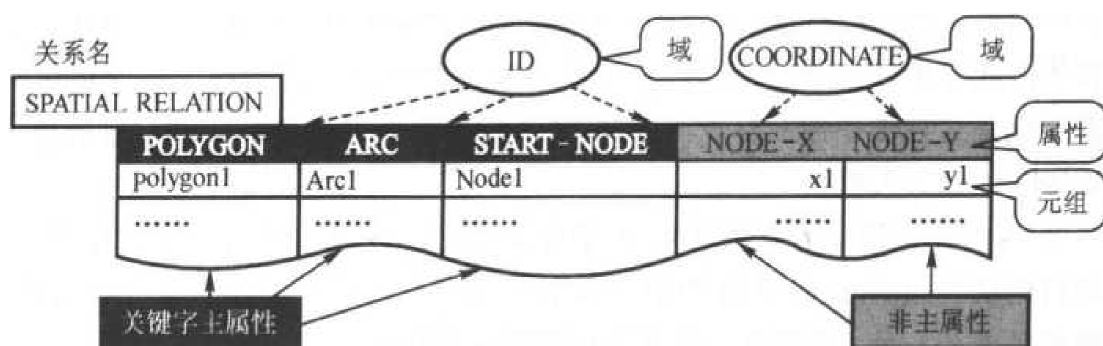
2.3 关系数据模型

关系模型是由 E. F. Codd 在 20 世纪 70 年代初首次引入到数据库领域中的。在关系模型中特别强调关系的表现形式与关系在计算机中的实现方法之间是相互独立的，即独立于数据在物理设备上的排列方式，独立于索引结构，独立于数据的存储路径，等等。因此，关系模型将用户对数据的视图与其最终的物理实现分割开来。从而，使得关系模型比层次模型和网状模型更有利于数据库的设计。

关系数据模型是一种数学化的模型，它把数据的逻辑结构归结为满足一定条件的二维表中的元素，这种表就称为关系。关系的集合就构成为关系模型。

1．关系数据模型的基本概念

(1) 关系。关系是一个二维表，表的每行对应一个元组，表的每列对应一个域。由于域可以相同，为了加以区分，对每一列起一个名字，称为属性。如图 4－15所示。



关系模式:SPATIAL-RELATION(POLYGON,ARC,START-NODE,NODE-X,NODE-Y)

图 4-15 关系模型的基本概念

(2) 关键字。关系中的某一属性组,若它的值惟一地标识了一个元组,则称该属性组为候选关键字。若一个关系中有多个候选关键字,则选定一个为主关键字。该关键字的诸属性称为主属性,其余的属性叫作非主属性。

(3) 关系模式。关系的描述称为关系模式。它包括关系名、属性名、属性向域的映射及属性间的依赖关系等。这里指的关系模式仅由前三部分组成,通常记为 $R(A_1,A_2,\cdots,A_n)$,R 为关系名,$A_1,A_2,\cdots,A_n$ 为属性名。属性向域的映射经常直接说明为属性的类型、长度。一个关系模式确定了这个关系的二维表的框架。

(4) 关系数据库。一系列关系的集合构成一个关系数据库。关系数据库有型、值之分。型即数据库的描述,它包括若干域的定义以及在这些域上定义的若干关系模式。值是这些关系模式在某一时刻的关系的集合。由于每个关系均有一个关系模式,因此构成关系数据库的所有关系构成了关系数据库模式。

从数据库设计的角度来理解上述概念时,可以认为数据库的设计主要体现了关系模式的设计。只有当数据装入以后,才能形成数据库。因此人们常把关系模式称为关系数据库的结构,而把关系称为关系模式的实例或外延。所谓空间数据库的设计,就是设计一组反映空间数据及其联系的合理的关系模式。

(5) 关系完整性。关系的完整性即指关系的正确性、相容性和有效性。它是给定的关系模型中数据及其联系的所有制约和依存规则,用以限定数据库状态及状态变化,从而保证数据的正确、相容和有效。

关系模型的完整性有三类:实体完整性、参照完整性和用户定义的完整性。其中,实体完整性和参照完整性是关系模型必须满足的完整性约束条件。

① 实体完整性。设属性 A 是基本关系 R 的主关键字组成部分(即主属性),则属性 A 不能取空值。关系数据模型中存在各种关系,如构成基本表的基本关系和导出关系(如查询表、视图表)等。实体完整性是针对基本关系的。一个基本关系通常对应现实世界的一个客体集合。例如,在地理空间中,“点”的基本关

系对应着地理客体(点、线、面和体)中的空间点客体集合。由于现实世界中的任何客体彼此是可以区分的,所以,它们具有某种惟一性标识。在关系模型中,主关键字就是这种惟一性的标识。因此,主关键字不能取空值。所谓空值是指"不知道"或"无意义"的值。

② 参照完整性。若基本关系 R 中含有与另一个基本关系 S 的主关键字 K 相对应的属性组 F(此时 F 称为 R 的外部关键字),则对于 R 中每个元组在 F 上的值必须或为空值,或等于 S 中某个元组的主关键值。

例如,一条弧段可以定义为具有起点(START-NODE #)、终点(END-NODE #)和左、右多边形标识(LEFT-POLYGON # ,RIGHT-POLYGON #)等等特征属性的基本关系 ARC,这条弧段可由系统确定一个弧段编号(ARC #)来惟一地标识。另外定义一个多边形基本关系 POLYGON,它可由多边形的内点坐标(*XC*, *YC*)和系统确定的一个多边形标识码 POLYGON # 等等特征属性组成。上面两个基本关系简化的模式分别如下:

ARC (*ARC # , START-NODE # 、 END-NODE # , LEFT-POLYGON # , ROGHT-POLYGON # , ...*)

POLYGON (*POLYGON # , XC, YC, ...*)

弧段关系 ARC 的主关键字为 ARC # ,多边形关系 POLYGON 的主关键字为 POLYGON # 。在 ARC 关系中 LEFT-POLYGON # 和 RIGHT-POLYGON # 不是关键字,同时这两个属性直接对应于关系 POLYGON 中的主关键字,所以成为关系 ARC 的外部关键字。这时 ARC 关系中的每个元组在 LEFT-POLYGON # 和 RIGHT-POLYGON # 上的取值只允许有两种可能:一是取空值,表明左右多边形尚未标识;一是取非空值,则 LEFT-POLYGON # 和 RIGHT-POLYGON # 的值必须是 POLYGON 关系中某些元组主关键字 POLYGON # 的值,即被参照的关系 POLYGON 中一定存在这样的元组,它的主关键字值等于关系 ARC 中的外部关键字值。这种值的对照就是参照完整性。

③ 用户定义的完整性。用户定义的完整性是指由应用环境决定,并针对某一具体数据库的约束条件必须满足的规则。所以,所谓用户定义的完整性,实则是反映了某一具体应用所涉及的数据必须满足的语义要求。

2. 空间数据库关系数据模型的概念设计和逻辑设计

空间数据库关系模式的构造,就是将点、线、面等空间客体的特征以关系模式加以表达和组织。这是空间数据关系模型逻辑设计过程中最重要的部分。关系数据库的规范化理论是关系数据模型逻辑设计的有力工具,也可用来指导空间关系数据模型的设计。主要可用在以下几方面:在数据分析阶段,用数据依赖的概念分析和表示各数据项之间的联系;在概念结构设计阶段,以规范化理论消除初步概念模型中的冗余联系;由概念模型向数据模型转换过程中,用模式分解

等设计方法指导设计。下面简要地介绍关系模型设计理论的三个主要部分：数据依赖、范式以及模式设计的方法等内容。

(1) 数据依赖。上述的数据完整性约束条件是对关系中值的一种限制条件。这种限制可分为两类：一是依赖于值域元素语义的限制，例如，地面的高程数据不会超过 10 000 m 等。这种约束的检查主要是由数据库管理系统的完整性子系统完成的，与数据库模式设计几乎无关。另一类是依赖于值的相等与否的限制，这类限制并不取决于某一元组的属性取什么值，而仅仅取决于两个元组的某些属性的值是否相等。这类限制统称为数据依赖，而函数依赖则是其中最重要也是最基本的一种数据依赖。

(2) 函数依赖。函数依赖是由数学理论派生的术语，它表征一个属性或属性集合的值对另一个属性或属性集合的值的依赖性。定义为：设 $R(U)$是属性集合 U 上的关系模式，X 与 Y 是 U 的子集，若对于 $R(U)$的任意一个当前值 r，如果对 r 中的任意两个元组 t 和 s，总有 $t[X]=s[X]$，就必有 $t[Y]=s[Y]$，则称 X 函数决定 Y，或 Y 函数依赖于 X。记作 $X\rightarrow Y$。也可形式化表示为：如果属性集合 X 中的每个属性的值构成的集合惟一地决定了属性集合 Y 中的每个属性的值构成的集合，则属性集合 Y 函数依赖于属性集合 X。属性集合 X 中的属性有时也称作函数依赖 $X\rightarrow Y$ 的决定因素。

需要强调的是，函数依赖是关系所表述信息本身具有的语义特性，而不能由属性构成关系的方式来决定，也不能由关系的当前内容所决定。

关系数据模型中的函数依赖主要有以下几种类型：

① 最基本的函数依赖是主关键字决定关系模式中的非关键字。如图 4－16 所示，由于每个空间客体在空间数据库中都有一个惟一的空间索引代码，所以空间索引代码 SPATIAL-INDEX 函数依赖于空间客体的系统标识码 SPATIAL-ID(主关键字)。从而有函数依赖：SPATIAL-ID →SPATIAL-INDEX。

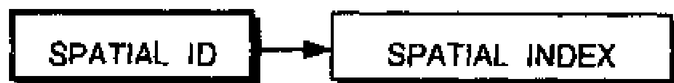


图 4－16　函数依赖

② 完全函数依赖。设有函数依赖 $X\rightarrow Y$，且对于任何 $X'\subset X$，$X'\rightarrow Y$ 都不成立，则称 $X\rightarrow Y$ 是一个完全函数依赖，记作 $X\Rightarrow Y$。反之，如果 $X'\rightarrow Y$ 成立，则称 $X\rightarrow Y$ 是部分函数依赖。

③ 传递函数依赖。设关系模式 $R(U)$，$X,Y,Z\subseteq U$，如果 $X\rightarrow Y$，$Y\rightarrow Z$，且 $XY\cap Z=\Phi$，$Y\rightarrow X$ 不成立，则有 $X\rightarrow Z$，称为 Z 传递依赖于 X。

④ 多值依赖。设 $R(U)$是属性集 U 上的一个关系模式。X,Y,Z 是 U 的子集，并且 $Z=U-X-Y$，如果对 $R(U)$的任一关系 r，对 r 中任意两个元组 t 和 s，

都有 $t[X]=s[X]$，就必然存在元组 w 和 v（这里 w 和 v 可以和元组 t 和 s 相同），使得 $w[X]=v[X]=t[X]=s[X]$，而 $w[Y]=t[Y]$，$w[Z]=s[Z]$，$v[Y]=s[Y]$，$v[Z]=t[Z]$，则称 Y 多值依赖于 X，记作 $X \to\to Y$。

(3) 范式。关系满足某种规范化的形式称为范式。由于关系的属性之间存在着多种多样的函数依赖特性，当用关系模式表达时，一些不良的语义特性就可能造成数据存储的冗余及由此造成的数据操作的异常，所以提出了关系模型必须遵循的一些规范化形式，简称范式。到目前为止，已提出了六种关系范式。其中 1NF、2NF、3NF 和 BCNF 属于函数依赖范畴，4NF 和 5NF 属多值依赖范畴。从一个低一级范式的关系模式，通过模式分解转换为若干个高一级关系模式集合的过程，叫做规范化。

① 第一范式(1NF)。是一个关系模式所要遵循的最基本的条件，即关系中的每个属性必须是原子的、不可分割的数据项。在关系数据库中，若一个关系模式满足 1NF，它就是合法的。同时 1NF 也是其他范式成立的前提条件。例如，图 4－3 所示的多边形与弧段的关系 POLYGON-ARC 可以用图 4－17(a)的关系模式表示。但不是 1NF，因为 ARCS 域不是简单域，它又包含了 ARC # 和 ARC-LENGTH 两个域。所以这种关系可以称为非规范化关系。将非规范化关系中的非简单域用简单域来代替，就可以将其转化为 1NF 的关系，这个过程就是规范化。如图 4－17(b)所示的即是多边形与弧段关系 POLYGON-ARC 转化为规范化的 1NF 所形成的关系模式。

POLYGON #	ARCS	
	ARC #	ARC-LENGTH
polygon1	arc1	100
	arc2	133
	arc3	150
polygon2	arc1	100
	arc4	166

图 4－17(a)　非第一范式关系表



POLYGON #	ARC #	ARC-LENGTH
polygn1	arc1	100
polygon1	arc2	133
polygon1	arc3	150
polygon2	arc1	100
polygon2	arc4	166

图 4－17(b)　第一范式关系表

② 第二范式(2NF)。是指关系在满足 1NF 的基础上，每一个非主属性完全

函数依赖于该关系的关键字。即当一个非关键字字段部分地函数依赖于关键字时，就违反了2NF。如图4-17(b)所示的关系虽然满足第一范式，但由于非主属性ARC-LENGTH并不是完全函数依赖于主关键字(即POLYGON#和ARC#的组合关键字)，而是部分依赖于主关键字，即完全函数依赖于主关键字的一个子集ARC#，所以该关系不是2NF。当一个关系模式不属于2NF时，就会产生数据的冗余存储和修改复杂化等问题。因此，在关系模式设计时需要将此关系分解为两个关系POLYGON#→ARC#和ARC#→ARC-LENGTH。这两个分解后的关系就满足2NF的要求。如图4-18所示。

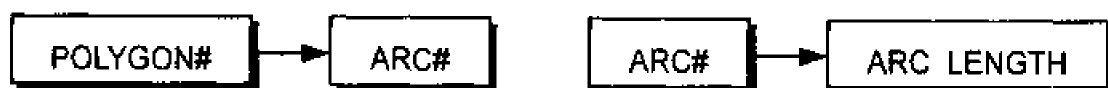


图4-18 分解为第二范式的关系表

③ 第三范式(3NF)。是指关系在满足第二范式的基础上，其非关键字属性(即非主属性)既非函数依赖，也不传递依赖于关键字，即属于3NF的关系，其非关键字属性之间都是独立的，不存在函数依赖关系。如图4-19(a)所示的关系符合2NF的要求，但不是3NF。因为非主属性"四叉树空间索引代码QUADTREE CODE"函数依赖于另一个非主属性"最小边界矩形MBR"，所以传递依赖于关系模式中的主属性POLYGON#。这种非3NF关系可以分解成两个3NF的关系。如图4-19(b)所示。

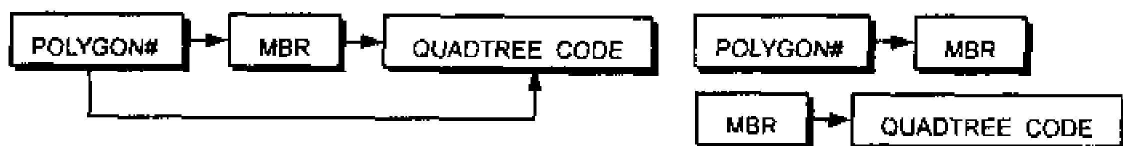


图4-19(a) 非第三范式关系中的传递依赖　　图4-19(b) 分解成两个第三范式关系

④ BC范式(BCNF)。是指如果关系模式R是第三范式，且没有一个非关键字属性是完全函数依赖于其他的非关键字属性，则关系模式R称为扩充第三范式的模式，即BC(Boyce Codd)范式模式。

规范化的基本思想是逐步消除数据依赖中的不合理部分，使模式中的各个关系达到某种程度的分离，让一个关系描述一个概念。这一个过程实质上是概念的单一化过程。

2.4 三种传统数据模型的比较

上述三种传统数据模型均可用来描述包括空间数据和属性数据的地理数据，但各有优缺点。如表4-1所示，在选用时应根据具体的应用要求择优选用，或取长补短。

表4-1 三种常用的传统数据模型比较

数据模型	优点	缺点
层次模型	1. 易于理解、更新与扩充 2. 通过关键字数据访问易于实现 3. 事先知道全部可能的查询结构，数据检取方便	1. 访问限于自上而下的路径，不够灵活 2. 大量索引文件需要维护 3. 一些属性值重复多次，导致数据冗余、存储和访问的开销增加
网络模型	1. 空间特征及其坐标数据易于连接 2. 在很复杂的拓扑结构中搜索，有环路指针是一种很有效的方法 3. 避免数据冗余，已有数据可充分使用	1. 间接的指针使数据库扩大，在复杂的系统中可能占据数据库的很大部分 2. 每次数据库变动，这些指针必须更新维护，其工作量相当大
关系模型	1. 结构灵活 2. 可以满足布尔逻辑和数学运算表达的各种查询需要 3. 允许对各种数据类型的搜索、组合和比较	1. 为找到满足指定关系要求的数据，许多操作涉及到对文件的顺序搜索，对大型系统而言，很费时间 2. 为保证以适宜速度进行搜索的能力，商用系统一般需经过十分精心的设计，故价格昂贵

§3 空间数据库概念模型设计——语义数据模型和面向对象数据模型

传统数据模型在数据库发展中起了重要的作用，但仍然存在许多的不足，可概括为：

(1) 以记录为基础的结构不能很好地面向用户和应用。传统数据模型基本结构是记录，而人类对现实世界的认识往往以某个事物或概念为单位。这些可以互相区别的事物或概念统称为客体。客体与记录不一定相对应。记录的组成往往从计算机实现的角度加以考虑，而不一定反映人们对现实世界的认识，不利于建立数据的概念模型。

(2) 不能以自然的方式表示客体之间的联系。除了对客体本身的描述之外，对客体间联系的描述也是数据模型的一个重要方面。层次和网状数据模型虽然提供了显式的联系描述手段，但这些联系形式不是客体间联系的自然表示，而是联系的物理实现的描述。这样就把本来应该对用户隐藏的物理实现的细节，当作数据模型的组成部分呈现在用户面前，既不自然又不便于用户理解和使用，而且也有损于数据的物理独立性。关系数据模型虽然避免了这个缺点，客体间的联系也可以用关系来表示，但关系数据模型很难从数据模型看出客体间的联系。一个客体与哪些客体有联系，必须检查每个关系才可能知道。现实世界中的客体联系被掩盖在记录、属性之中。总之，三种传统数据模型都难以自然地

表示客体间的联系,这是它们的一个主要弱点。

(3) 语义贫乏。三种传统数据模型都或多或少含有一些语义,例如关系数据模型的主关键字,层次数据模型中的主从关系,网状数据模型中的系都提供了一定的语义信息。但是就总体而言,它们能够表达的语义是贫乏的。记录各个属性间的关系及其作用不够明确,这就要求用户通过文字来描述这些记录的语义,但这些文字描述又不是数据库的一部分。此外,它们对客体间联系的语义描述,也缺少必要的手段。"交互关系"、"从属关系"、"性质"、"组成"等都是代表语义不同的联系,却用同样的记录结构表示,并且缺少描述这些联系的相应结构,在语义上无法区分这些联系。

(4) 数据类型太少难以满足应用需要。传统数据模型原来都是面向事物处理的。它们一般只提供最常用的一些简单数据类型,例如整数、实数、字符串等,而且不允许用户自己定义新的数据类型。但在有些应用中,希望属性值不直接给出,而是由公式、规则或过程推导而得。有时不但要知道数据的当前值,还要知道不同时间的数值。这些功能在传统数据模型中都不支持。

为了克服传统数据模型存在的缺陷,从 20 世纪 70 年代后期开始,陆续出现了一些其他的数据模型,其中最主要的是语义数据模型和面向对象数据模型。

3.1 语义数据模型

语义模型的模型结构是由若干种抽象所组成,用这些抽象来描述客体的基本语义特性,再根据语义模型结构规则把这些抽象有机地组织起来。语义模型的抽象是比数据元素更高的层次,通过抽象在客体层次上构造模型,形成客体类别及这些客体类别之间的关联,在基本语义结构被确定后,再加入数据元素。不过,由于语义模型不具有避免异常和消除冗余的准则,所以需要与关系模型在设计的不同阶段相互配合,以求达到最佳的设计效果。

最常用的语义模型之一是实体联系模型(entity-relationship model,E-R 模型)。E-R 模型为数据库分析设计人员提供了三种主要的语义概念,即实体、联系和属性。设计人员可以通过一种称为 E-R 图的图示方法来组织 E-R 模型的设计。

(1) 实体。实体是对客观存在的起独立作用的客体的一种抽象。在 E-R 图中,用矩形符号表示实体。实体的命名标注于矩形符号之内。如图 4-20 所示。

(2) 联系。联系是客体间有意义的相互作用或对应关系。一般可分为一对一的联系(1:1),一对多的联系($1:N$)和多对多的联系($M:N$)三种类型。联系在 E-R 图中用菱形符号表示,联系的名称同样标注于菱形之中。实体和联系之间用线段连接,并在线上注明连接的类型。

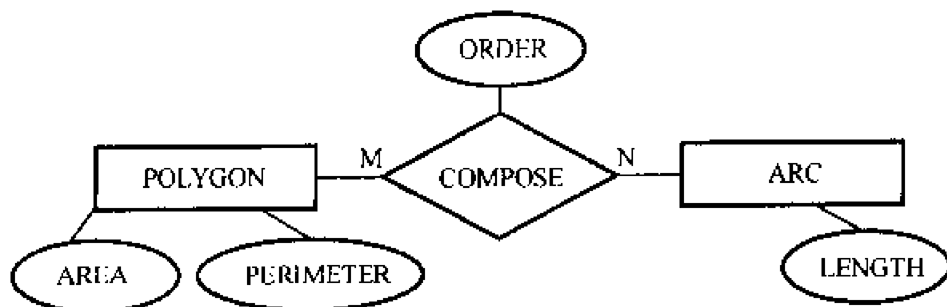


图4-20　E-R模型的图形表示

(3) 属性。属性是对实体和联系特征的描述。每个属性都关联一个域(值的集合)。属性用一个椭圆形表示,椭圆中放置属性的名称,属性同实体和联系之间也用线段连接。

采用E-R模型进行数据库的概念设计可以分为以下三个步骤:

(1) 设计局部的E-R模型。在局部E-R模型的设计中,主要需要完成局部结构范围的确定,定义属性、实体和联系,以及属性的分配等。如图4-21所示。

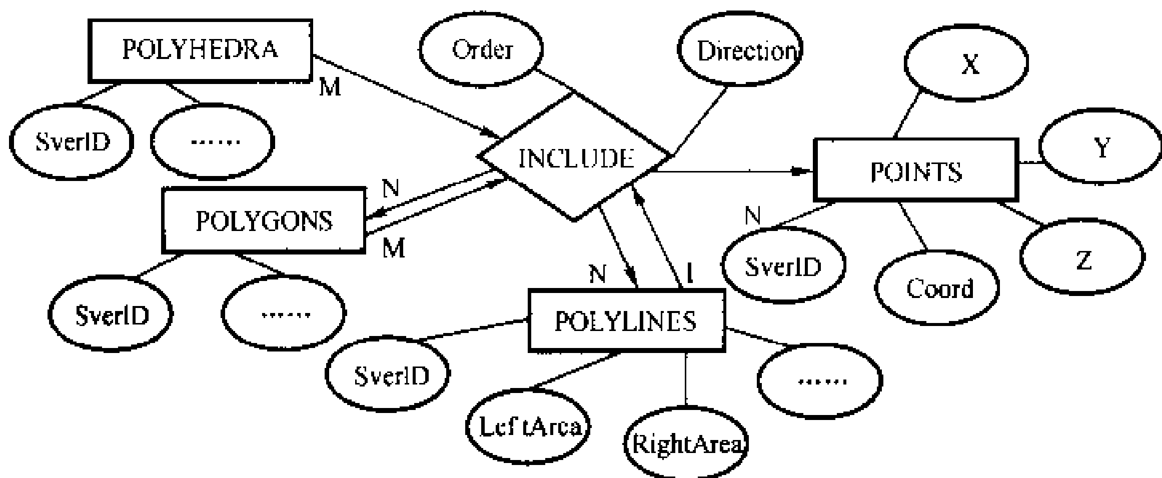


图4-21　简化的空间数据局部E-R模型

(2) 设计全局的E-R模型。这一步骤是将所有局部的E-R图综合成单一的全局的E-R图,即全局的概念模型。设计全局的概念模型的过程中,首先需要确定公共的实体类型,然后反复进行两个局部E-R图的合并,检查并消除冲突,直到所有的局部E-R模型都合并成一个完整的全局E-R模型。其中,当合并两个E-R图时,可能会遇到三类冲突,即:属性冲突,包括类型、取值范围、取值单位的冲突;结构冲突,如作为实体又作为联系或属性,同一实体其属性成分不同等;命名冲突,包括实体类型名、联系类型名之间异名同义,或异义同名等。

(3) 全局E-R模型的优化。一个好的全局E-R模型除能反映用户需求之外,还应满足下列条件:实体类型个数尽可能少,实体类型所含的属性尽可能少,实体类型间联系无冗余。优化就是要达到这三个目的,即通过实体类型的合

并,指相关类型的合并来实现。一般把 1:1 联系的两个实体类型合并,具有相同关键字的实体类型也可以合并成一个实体类型,但要考虑空值存储和速度问题;还有冗余属性的消除,要注意效率,可存在适当的冗余,这要根据具体情况而定;最后是冗余联系的消除,冗余的联系应该从 E-R 图中删除。

E-R 模型具有一些明显的优点,即接近人的思想,易于理解,同时又与计算机具体的实现无关,是一种很好的数据库概念设计的方法。然而,正是由于它与计算机的具体实现无关,所以没有一个数据库管理系统直接支持 E-R 模型的实现。这是因为 E-R 模型只说明了实体间的语义上的联系,而没有进一步地详细阐明实体的数据结构,不能在数据库中直接实现。所以一般用 E-R 模型设计数据库的概念模型,然后在逻辑设计阶段,把 E-R 模型转换成计算机能够实现的数据模型,比如关系数据模型、层次数据模型或网状数据模型等。

3.2 面向对象的数据模型

对象与实体一样是客观世界中客体的一种抽象的描述,它由客体的数据和对数据的操作组合而成。这是面向对象方法的基本概念之一。此外,类是对多个对象共同特性的抽象概括。消息是对象之间通信的惟一方式,用来指示接收消息对象的操作。方法是对象收到消息后应采取的动作序列的描述。对象的实例则是指由一特定类描述的具体对象的实现。对象具有许多特性,如封装性、继承性、多态性等等,并可通过数据的持久性支持数据库的功能。另外,语义数据模型中的概括、聚合、派生等概念也被面向对象的方法所采纳。当然,传统的数据库中有关存储管理、并发、数据恢复和查询语言等的概念,也必须包含在数据模型中,这样,就形成了面向对象的数据模型。

1. 面向对象数据模型的基本概念

(1) 对象。一个对象就是现实世界中一个客体的模型化,与数据库中记录、元组等概念相似,但更为复杂。它具有一个惟一的名称标识,并且把自身的状态和内在的功能封装在一起。在面向对象的数据模型中,一个对象的状态是通过域来描述的,可称为私有存储单元。对象的封装性特征在此处表现为它的私有存储单元只能由它自己的功能进行处理。对象是对现实世界的一种高度的抽象概括形式。在地理信息系统的空间数据库中,任何一个空间上的客体都可以用对象的形式加以表达,比如表示一个行政区域的多边形对象、表示一条河流的弧段对象等等。

(2) 消息。消息是对象之间相互请求或相互协作的惟一途径。一个对象必须通过向其他对象发送消息的形式使得其他对象提供各自所能实现的功能。在面向对象的数据模型中,消息是对象之间惟一的通信形式,也就是外界能够引用对象操作及获取对象状态的惟一方式。这个特性保证了对象的实现只能依赖于

它本身的状态和所能接受的消息，而不依赖于其他对象的内部细节和状态。例如，把一个空间数据库以一个对象的形式进行实现，则空间数据库中所有操作功能的运作都必须是响应数据库用户传送给他的消息的结果。

消息有公有和私有之分，公有和私有消息的含义是：如果一些消息都属于同一个对象，其中有些是可由其他对象向它发送的，叫做公有消息。另外一些则是由它自己向自身发送的，就叫做私有消息。空间数据库的面向对象模型中，一个复杂对象，比如多边形对象等，往往是由一些简单对象组成的，比如弧段对象等。多边形对象中实现了许多消息，一些是直接针对多边形对象本身的，比如要求计算出多边形的周长等，可由外界对象发送给它，这些就是公有消息。另外一些消息是针对多边形对象所包含的更简单对象的，比如多边形对象要求组成它的各个弧段对象分别计算出各自的长度，这就需要给这些弧段对象发送消息。这些消息是不对外开放的，外界也不必了解它，这些消息就可当作私有消息来实现。如图4-22所示。

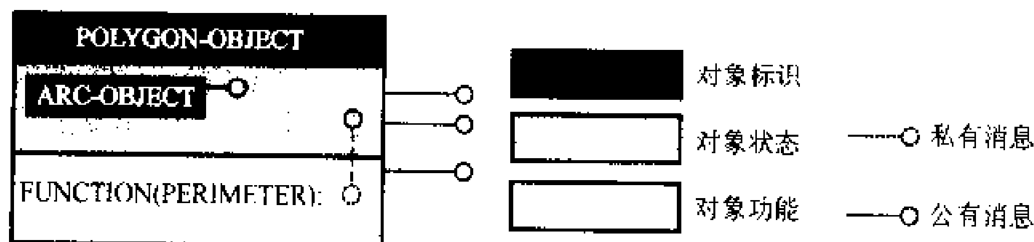


图4-22　对象的组成(多边形对象)

(3) 类。是对一组对象的抽象描述，它将该组对象所具有的共同特征集中起来，以说明该组对象的能力和性质。实例是某类的一个具体对象，类是多个实例的抽象综合。可见类和实例之间是抽象和具体的关系。类的确定方法是归纳，这就要求在观察问题时，要注意提炼所遇到对象的共同特征，以确定其类的表达。

2. 继承及类之间的层次关系

(1) 继承是现实世界中对象之间的一种独特关系，它使得某类对象可以自然地拥有另外一类对象的某些特征和功能。继承性可以极大地简化数据模型的设计，因为继承性使许多对象的特性和功能可以相互依赖，所以对象之间某些相同或相似的特征和功能就不需要重复地实现，只要通过继承就可以相互借用和共享。可见继承性具有双重作用，一是减少代码的冗余，二是通过协调性简化对象类相互之间的接口和界面。

继承的分类从对象类的数量上看可分为单继承和多继承两种。例如直线段对象类继承了曲线对象类的特性和功能，可以直接从曲线对象类中派生出来，则这种继承就是单继承类型。如图4-23(a)所示。而多边形对象类既可以继承

曲面对象类的特性，其边界又具有曲线对象类的特性，所以可以从上述两个对象类中共同继承而来，这就是多继承类型。如图 4－23(b)所示。

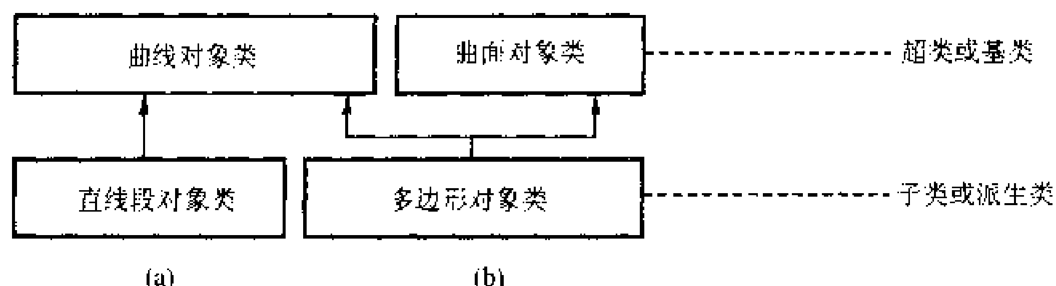


图 4－23 对象类的单继承和多继承及其层次结构

(2) 类的层次结构是引入类的继承关系而产生的。被继承的类称为超类或基类，继承其他类的对象类称为子类或派生类。类的层次如图 4－23 所示。超类和子类的概念是相对的。类的层次结构是用于描述和组织世界事物的一种简单、有效的工具。

3. 功能重载和多态性

面向对象系统中消息传递的实现是以重载和多态为支撑概念的，功能重载实际上意味着实现特定功能的方法不仅以名称来区分，而且用它所带的参数来区别。例如在地理信息系统的图形显示子系统中需要显示一个多边形的区域，系统可能会提供两个不同的对象绘图方法。其一是从一个数据库系统中提取多边形数据来绘图，另一个是从一个图形文件中提取多边形数据来绘图。第一种方式需要将数据库的一个连接作为参数传递给对象方法；第二种方式则需要将图形文件的路径作为参数传递给对象方法。这两种对象方法虽然其名称可能采取相同的写法，但由于所调用的参数的差异，面向对象的系统就会根据功能重载的原则将两个对象方法加以区分。如图 4－24(a)所示。

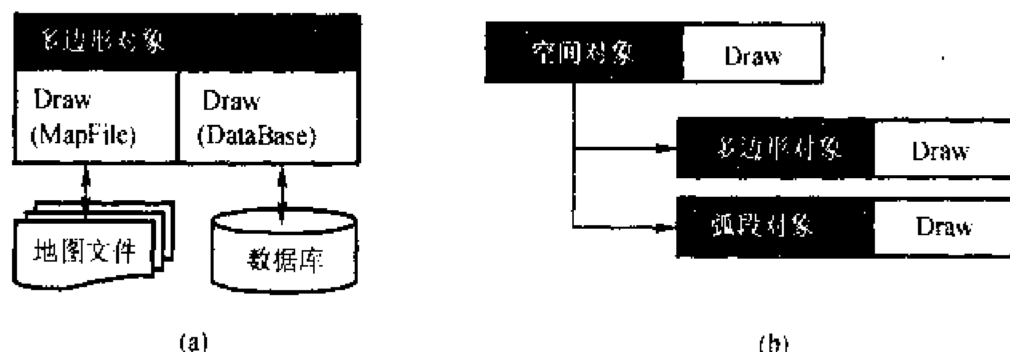


图 4－24 面向对象的功能重载和多态

多态的含义是同一个消息可以根据发送消息对象的不同采用多种不同的行为方式。例如在上述的图形系统中，有多边形、弧段、点等几类对象，在这些图形对象类的超类中可以定义一个虚拟的绘图方法 Draw。而在具体的多边形、弧段

和点类中再分别定义各自特定的绘图方法 Draw。这样,面向对象的系统在绘制某一区域的所有空间对象的图形时,就可以对所有的图形对象发送同一种对象绘图消息,而不同的对象以各自特定的方式响应同一种绘图消息。如图 4-24(b)所示。

4. 概括和聚集

(1) 概括。概括的含义是把一组具有相同特征和操作的对象类归纳在一个更一般的超类中。例如,多边形对象类是一种特殊的空间对象,而弧段对象类也是一种特殊的空间对象,可见,这里的空间对象类归纳了多边形对象类和弧段对象类共同具有的一些空间特征,相比之下是更为一般意义上的对象类。所以,多边形对象类和弧段对象类可作为空间对象类的子类,而空间对象类则是它们的超类。概括形成的对象类不要求一定是现实中具体存在的对象类,它可以是一种抽象的对象类。从本质上看,概括形成了子类和超类之间一种称为 is-a 的语义联系,如图 4-25(a)所示。

(2) 聚集。聚集反映了嵌套对象的概念,嵌套对象是由一些其他对象组成的,它是用来描述更高层次对象的一种形式。例如,上述的地理信息系统图形显示子系统中,一个图层对象类是由多边形对象类、弧段对象类、点对象类等等对

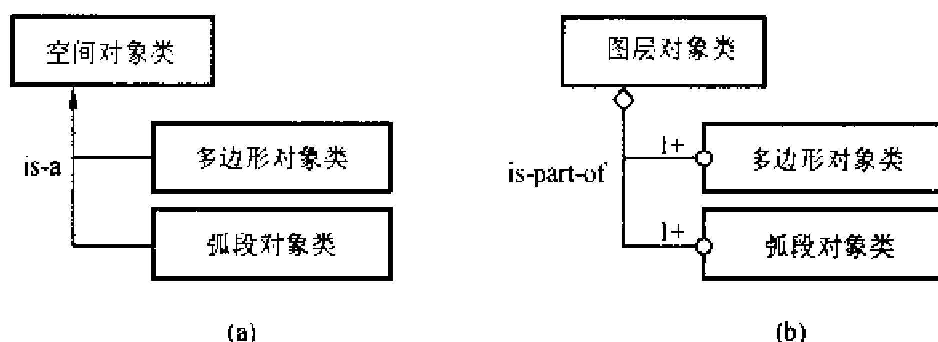


图 4-25　对象类的概括与聚集

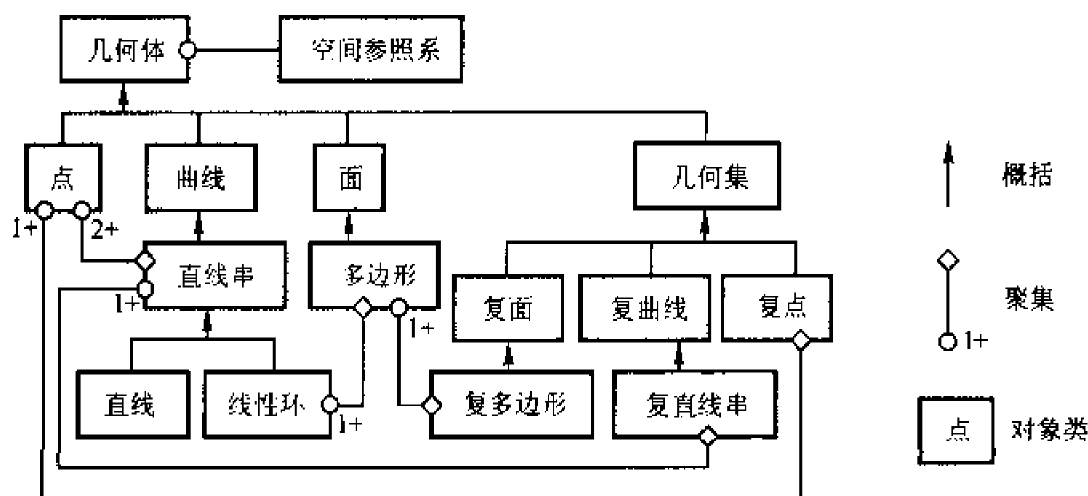


图 4-26　简单几何特征集关系

象类的聚集体。这里的图层对象就是嵌套对象,多边形对象类等与图层对象之间形成一种 is-part-of 的语义联系,如图 4-25(b)所示。

5. 空间数据库对象模型

运用上述面向对象的概念和方法,就可以建立起地理信息系统中空间数据库的对象数据模型。OpenGIS 协会推荐了一种空间对象数据模型,其中的简单几何特征集如图 4-26 所示。该模型以面向对象的方式将常见的一些空间几何对象按概括和聚集关系组织成对象层次。

§4　空间数据库逻辑模型设计和物理设计

逻辑设计的目的是从概念模型导出特定的数据库管理系统可以处理的数据库的逻辑结构(数据库的模式和外模式),这些模式在功能、性能、完整性和一致性约束及数据库可扩充性等方面均应满足用户提出的要求。

4.1　空间数据库逻辑设计的步骤和内容

(1) 初始模式形成。把 E-R 图表示的的实体-联系类型,转换成选定的数据库管理系统所支持的记录类型,包括层次、网状、关系、面向对象系统等。

(2) 子模式设计。子模式是应用程序与数据库的接口,允许有效访问数据库而不破坏数据库的安全性。

(3) 模式评价。对逻辑数据库结构即模型,根据定量分析和性能测算作出评价。定量分析是指处理频率和数据容量及其增长情况。性能测算是指逻辑记录访问数目、一个应用程序传输的总字节数和数据库的总字节数等。

(4) 优化模式。为使模式适应信息的不同表示,可利用数据库管理系统性能,如建索引、散列功能等,但不修改数据库的信息。

4.2　E-R 模型向关系数据模型的转换

E-R 模型可以向现有的各种数据库模型转换,对不同的数据库模型有不同的转换规则。例如,需要实现一个全关系型的空间数据库,则向关系模型转换的规则是:

(1) 一个实体类型转换成一个关系模式,实体的属性就是关系的属性,实体的关键字就是关系的关键字。

(2) 一个联系类型转换成一个关系模式,参与该联系类型的各实体的关键字以及联系的属性转换成关系的属性。该关系的关键字有三种可能情况:① 若

联系的类型为1:1,则每个实体的关键字均是该关系的候选关键字;② 若联系的类型为1:N,则关系的关键字为N端实体的关键字;③ 若联系的类型为M:N,则关系的关键字为各实体的关键字的组合。

如图4-21所示的简化空间数据局部E-R模型,可以照此规则转换成关系数据库的模式,如图4-27所示。

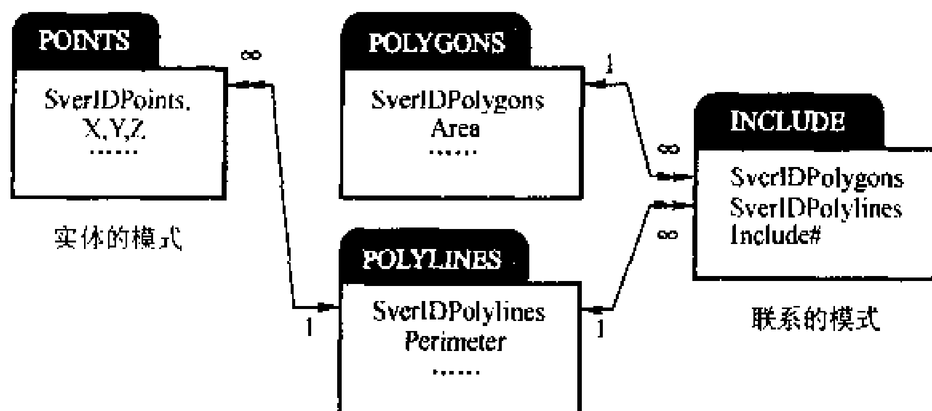


图4-27 E-R模型向关系数据库模式的转换

4.3 关系数据库的逻辑设计

关系数据库的逻辑设计过程包括以下步骤:① 导出初始关系模式。将E-R图按规则转换成关系模式。② 规范化处理。该步骤的任务是消除异常,改善完整性、一致性和存储效率。一般使关系达到3NF就可以满足使用。规范化过程实际上就是单一化过程,即一个关系描述一个概念。如果关系中多于一个概念,就把它分离出来。③ 模式评价。目的是检查数据库模式是否满足用户的要求,包括功能评价和性能评价。④ 优化模式。如果模式有疏漏,就要新增关系或属性;如性能不好,则需要采用合并、分解或选用另外的结构等手段来优化模式。其中:合并是对具有相同关键字的关系模式,如果对它们的处理主要是查询操作,且常在一起使用,可将这类关系模式合并;分解是指关系虽已达到规范化的要求,但如果某些关系的属性过多,则会影响效率,可将它分割成为两个或多个关系模式,按属性组分解的称为垂直分解,垂直分解需满足得到的每一个关系都包含主关键字。⑤ 形成数据库的逻辑设计说明书。逻辑设计说明书包括:模式及子模式的集合,可用数据库管理系统的语言描述,也可列表描述;应用设计指南,设计访问方式、查询路径、处理要求、约束条件等;物理设计指南,包括数据访问量、传输量、存储量、递增量等。

4.4 空间数据库的物理设计

空间数据库的物理设计,是从一个满足用户信息需求的、已确定的逻辑数据

库结构(即逻辑模型)出发,研制出一个有效的、可实现的物理数据库结构(存储结构或物理模型)的过程。物理设计常常包括某些操作约束,如响应时间、存储要求等。

物理设计可分为五步,前三步为结构设计,后两步为约束和程序设计。各步的具体内容是:

(1) 存储记录的格式设计。对数据项类型特征作分析,对存储记录进行格式化,决策如何进行数据压缩或代码化。使用"记录的垂直分解"方法,对含有较多属性的关系,按其中属性的使用频率不同进行分割;或使用"记录的水平分解"方法,对含有较多记录的关系,按某些条件进行分割,并把它们定义在相同或不同类型的物理设备上,或在同一设备的不同区域上,从而使访问数据库的代价最小,提高数据库的性能。

(2) 存储方法设计。物理设计中最重要的一个考虑是把存储记录在全数据库范围内进行物理存储安排。存储的方法主要有:① 顺序存储,该存储方式的平均查询次数为关系记录个数的 1/2;② 散列存储,该存储方式的查询次数由散列算法所决定;③ 索引存储,该存储方式需要确定创建何种索引,及在哪些库和属性上建立索引;④ 聚簇存储,"记录聚簇"是指将不同类型的记录分配到相同的物理区域中去,充分利用物理顺序性的优点,提高访问速度,即将经常在一起使用的记录聚簇在一起,以减少物理输入/输出的次数。

(3) 访问方法设计。访问方法设计为存储在物理设备上的数据提供存储结构和查询路径,该设计与选用的数据库管理系统有很大关系。

(4) 完整性和安全性考虑。根据逻辑设计说明书中提供的对数据库的约束条件、具体选择的数据库管理系统和操作系统的性能特征及硬件环境,设计建立数据库完整性和安全性措施。

(5) 应用设计。该设计包括人机界面的设计、输入/输出格式的设计、代码设计、处理加工设计等。

(6) 形成物理设计说明书。物理设计的结果是物理设计说明书。包括存储记录格式,存储记录位置分布及访问方法、它能满足的操作需求,并给出对硬件和软件系统的约束。在设计过程中,效率问题的考虑只能在各种约束得到满足且获得可行方案之后进行。此外,物理设计中应充分注意物理数据的独立性,所谓物理数据的独立性是指消除由于物理数据结构设计的改变而引起对数据库应用程序的修改。

§5 GIS空间时态数据库

5.1 空间时态数据库概述

任何事物的发生或演变都有其时间特性。从地理信息系统的角度来观察，时间在逻辑上可以是一条没有端点、向过去和未来无限延伸的坐标轴——时间轴，在每一设定的时间分辨率的坐标点上，都可以扩展其三维空间数据。它是现实世界的第四维，除了与三维空间一样具有通用性、连续性和可量测性外，还具有运动的不可逆性或称单向性。时间和空间不可分割地联系在一起。在一般的空间数据应用中，一些与时间关系极其密切的如环境监测、地籍管理等应用，时间应该是不可缺少的一维。它不仅仅作为数据的一个组成部分，而且与空间数据相互关联地存在着。

一个空间对象，除了空间和属性描述信息外，还要考虑时间信息。考虑对象从产生到消亡的时间信息，即对象在现实世界中存在的时间区间，一般称为有效时间或生命周期。若理论模型允许目标消亡后可以再生，则有效时间可以是多个不相交的时间区间的合并。另一方面，对象数据输入系统的时间称为执行时间或事务时间(transaction time)，也称系统时间(system time)。当然，用户还可以根据需要为对象标注时间，称为用户定义时间(user-defined time)。

虽然某些地理现象或地理事件并不一定需要记录和存储其时间数据，但如果地理信息系统具有支持时间的功能，能将时间变量加入到地理信息系统空间分析过程中去，则将具有更大的应用范围。因此，时空复合操作，即将空间分析的问题进一步拓展为时空分析的范畴，已成为地理信息系统的重要组成部分。时空地理信息系统是一种四维(X,Y,Z,T)或(S,T)的信息系统，其中(X、Y、Z)或(S)表示空间系统，(T)表示时间，这是一种具有时空复合分析功能和多维信息可视化的系统。

时空数据库系统或数据处理技术，其内容表现为以下三个主要方面：

(1) 空间时态数据的表达。空间时态数据表达的目的在于建立空间时态一体化数据模型。它涉及到时间标志、空间时态版本的标识、空间变化类型的定义、空间拓扑与时态拓扑、空间时态数据的存储结构，以及存取策略等内容。

(2) 空间时态数据的更新。空间时态数据的更新研究空间数据更新的类型、操作方法，更新对时空数据库中空间和时态拓扑的影响，以及拓扑重建等问题。

(3) 空间时态数据的查询。空间时态数据的查询探讨空间时态数据的各种跟踪算法，多维信息的复合、分析、可视化等。

5.2　时空一体化数据模型

地理信息系统中时空一体化数据结构和数据模型注重以一体化的形式表达地理数据在时间和空间方面的特性。这些模型主要有以下几种：

1. 时间片快照模型

常规的时空数据模型(如遥感图像数据的采集方式)即属于时间片快照(time-slice snapshots)模型。它是用一系列状态对应的地理数据来反映地理现象的时空演化过程。其中可分为矢量快照模型和栅格快照模型两类。这种模型的好处就是它可以直接在当前的地理信息系统软件中实现。另一个优点就是当前的数据库总是处于有效状态。但这种快照图像仅代表地理现象的瞬时状态，而缺乏对现象所包含的对象变化的明确表现，因此它不能确定地理现象所包含的对象之间在时间上的拓扑联系。在查询跟踪能力方面，由于没有时间维拓扑联系的存在，这一模型将无法实施按时间联系而确立的查询跟踪规则。此外，由于时间片快照是对状态数据的完整存储，所以造成数据的冗余度极大。如图4-28所示。

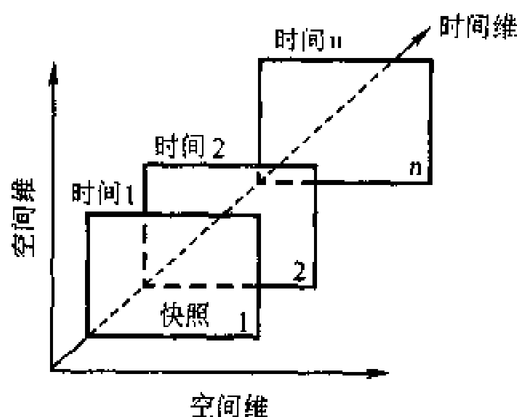


图4-28　时间片快照模型

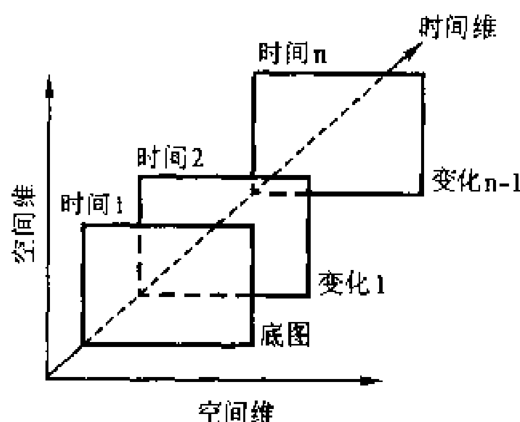


图4-29　底图叠加模型

2. 底图叠加模型

底图叠加(base map with overlay)模型由于类似于地理底图的修订方式而得名。它的基本思路是，首先确定空间数据的初始状态，即底图数据。然后按照适宜的时间间隔记录数据随时间发生的变化。再通过空间叠加操作，利用记录的变化数据来恢复各个时间片的状态数据，每一次叠加则表示状态的一次变化。如图4-29所示。

3.时空合成模型

时空合成(space-time composites)模型是在底图叠加模型的基础上提出的，其设计思想是将每一次独立的叠加操作转换为一次性的合成叠加。这样，变化的累积即形成最小变化单元，由这些最小变化单元构成的图形文件和记录变化历

史的属性文件联系在一起,则可以较完整地表达数据的时空特征。如图4－30所示。这一数据模型一方面保持了底图叠加模型的优点,另一方面与地理信息系统数据库表达空间、属性数据的手段相似,因而便于以矢量为基础的地理信息系统软件的实现。

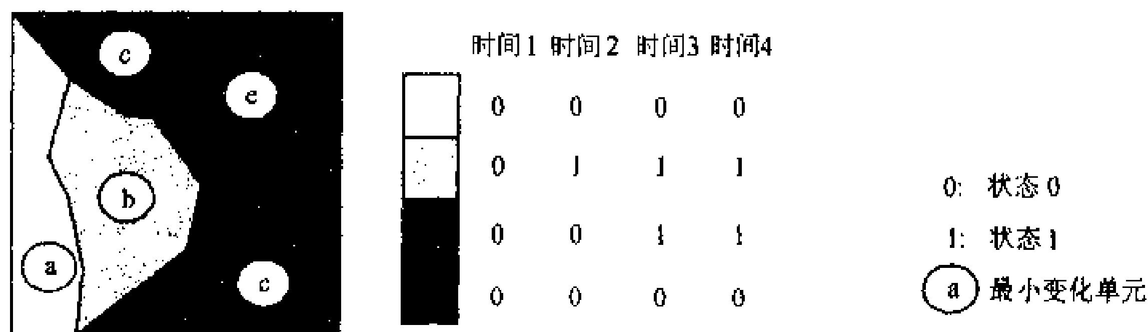


图4－30　时空合成模型

4. 全信息对象模型

全信息对象指包含空间、时态和属性信息的地理对象。全信息对象模型即是运用面向对象的设计方法,将地理现象和过程的空间与属性信息随时间的变化封装成由时态版本组成的对象,即全信息对象。模型中需要定义一个全信息对象的标识,由该标识惟一地确定各个地理对象。而每个全信息地理对象定义成随时间变化的多个地理空间属性实体的时态版本组合序列。其中既包含了空间数据、空间拓扑数据、时态数据、时态拓扑数据,同时也包含属性数据等。全信息对象模型是地理信息系统空间数据库一体化表达的方案之一。

在全信息对象模型中一个关键性的问题在于对象的更新模式,即空间和属性数据随时间的变化形式。更新模式直接关系到全信息对象的定义。从本质上看,更新模式主要有两种,一是对象的版本进化,二是新对象的产生。版本进化并不影响全信息对象的对象标识,仅仅是递增新的版本。而新对象的产生却会在数据库中引入新的对象标识。其实就对象版本进化与新生对象这两种最为基本的数据更新情况而言,最终是与具体应用领域密切相关的。无论是空间位置的变化、空间拓扑关系的变化,还是属性的变化,究竟哪些可以造成对象的版本进化,而哪些造成旧对象被废弃,新对象产生,都必须由最终的空间数据库用户在更新操作发生时根据具体的应用加以确定。所以系统的全信息对象标识的确定必须取决于特定用户的需求,而不是由系统的开发者所能决定的。

不同的地理对象定义将产生不同数据组织基础的概念模型。全信息地理对象定义综合了空间时态与属性,由此产生的概念模型则可同时满足基于特征和基于事件的要求。其中,为满足基于事件的概念模型的设计,就要求在模型中保存一串事件列表的数据。该列表以时间先后顺序表达所有的地理事件,每个地

理事件具有一个惟一的事件编码和时间标志，以及对事件的描述信息和与该事件相关的所有发生变化的空间、时态属性对象的连接等。如图4-31所示。

要设计满足基于特征的模型，每个全信息对象需要包含多个时态版本。时态版本以一个版本标识来确定。各个时态版本之间需要建立反映前后时序关系的时态拓扑，并以此建立起各个全信息对象的时间序列。同时，每个时态版本还可建立反映空间变化和属性变化的联系，用以连接同一事件中发生变化的其他对象的时态版本，以满足基于事件设计的需求，如图4-31所示。

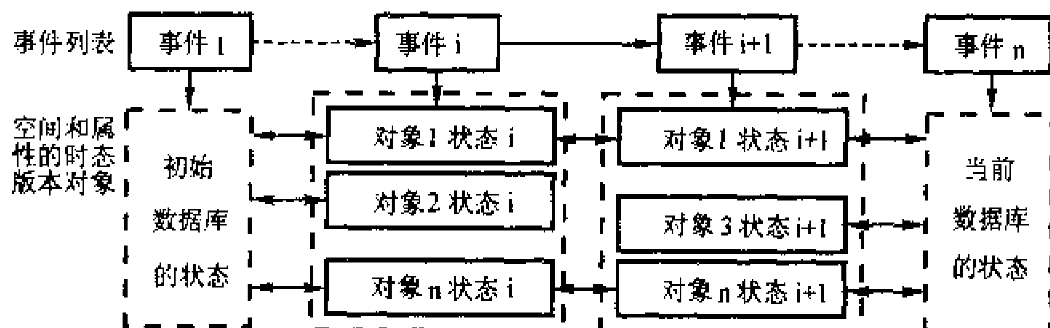


图4-31　全信息对象模型

将全信息对象模型以实体-联系方法表示为数据库的概念模型，如图4-32所示。

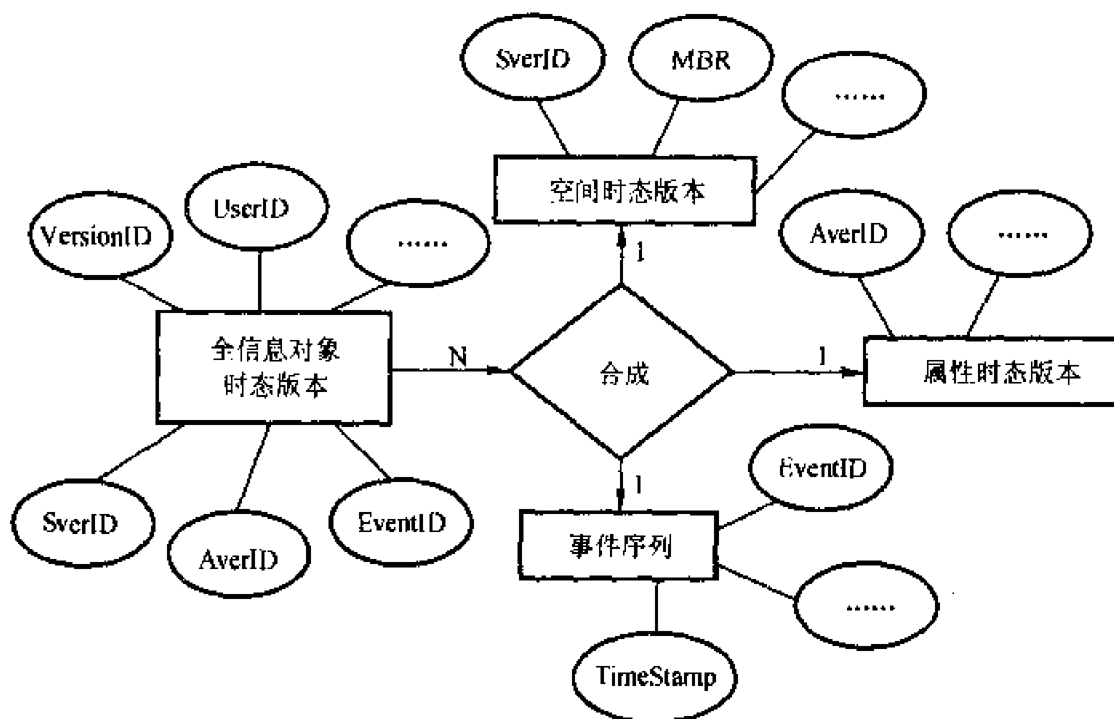


图4-32　全信息对象模型的局部实体-联系模型

思考题

1. 空间数据库的概念及其组成部分有哪些？

2. 空间数据库的设计有哪些主要的步骤和内容？

3. 传统的数据库模型有哪三种？它们各自的优缺点是什么？

4. 语义模型中的实体－联系模型的图示中有哪些基本元素？

5. 什么是关系数据模型？其中的基本概念如：元组、域、属性、关键字、关系模式、关系数据库、关系完整性等的含义分别是什么？

6. 在关系数据模型设计中，函数依赖和范式的含义是什么？对于下列关系模式

ARC（ARC #，START-NODE #，END-NODE #，LEFT-POLYGON #，AREA-LEFTPOLYGON，RIGHT-POLYGON #，AREA-RIGHT-POLYGON）；

判断是否符合范式的要求，并提出分解模式。其中，ARC # 为弧段标识，START-NODE #，END-NODE # 为弧段起止节点号，LEFT-POLYGON #，RIGHT-POLYGON # 为左右多边形标识，AREA-LEFTPOLYGON，AREA-RIGHT-POLYGON 为左右多边形的面积。

7. 面向对象数据模型中的基本概念如：对象、消息、类、继承、重载、多态、概括、聚集等的含义是什么？

第5章 空间分析的原理与方法

空间分析是地理信息系统科学内容的重要组成部分，也是评价一个地理信息系统功能的主要指标之一。早期的地理信息系统由于空间分析功能比较弱，常常引起与机助制图(CAC)和机助设计(CAD)之间的混淆，因为无论 GIS、CAC 或 CAD 都需要图形数字化和自动制图。但是 GIS 的目的，不但需要自动制图，它更主要的目的是为了分析空间数据，提供空间决策信息。因此，空间分析是地理信息系统区别于其他类型系统的一个最主要的功能特征，也是各类综合性地学分析模型的基础或构件。

关于空间分析，不少学者曾进行过研究，如 Haining 的《社会与环境科学中的空间数据分析》(1990)、Unwin 的《空间分析入门》(1981)和 Goodchild 等的《GIS 环境下的空间分析》(1994)等。Goodchild 对空间分析的框架作了较为系统的研究，他将空间分析分为两大类，即"产生式分析"(product mode)和"咨询式分析"(query mode)。前者指通过分析可以获取新的信息，尤其是综合信息；后者旨在回答用户所提出的问题。

空间分析的根本目的，在于通过对空间数据的深加工或分析，获取新的信息。因此，关于空间分析的定义可以这样表述：空间分析是基于空间数据的分析技术，它以地学原理为依托，通过分析算法，从空间数据中获取有关地理对象的空间位置、空间分布、空间形态、空间形成、空间演变等信息。

按照 Goodchild 提出的空间分析框架，可以将空间分析方法分为以下两种类型：① 产生式分析：数字地面模型分析，空间叠合分析，缓冲区分析，空间网络分析，空间统计分析；② 咨询式分析：空间集合分析，空间数据查询。

§1 数字地面模型分析

数字地面模型(Digital Terrain Model，简称 DTM)是定义于二维区域上的一个有限项的向量序列，它以离散分布的平面点来模拟连续分布的地形。按平面上等间距规则采样，或内插所建立的数字地面模型，称为基于栅格的数字地面模型，可以写成以下形式：

$$\mathrm{DTM} = \{Z_{i,j}\},\ i = 1,2,3,\cdots,m-1,m;$$

$$j=1,2,3,\cdots,n-1,n。$$

式中，Z 为栅格结点(i,j)上的地面属性数据，包括土地权属、土壤类型、土地利用等。当该属性为海拔高程时，则该模型即为数字高程模型(详见第3章§5)，它是各种地学分析、工程设计和辅助决策的重要基础性数据，有着广泛的应用领域，例如在地学分析中，用于自动提取各种地形因子，制作地形剖面图和划分地表形态类型；在工程勘测和设计中，可用于各种线路的自动选线、库坝的选址，以及土方、库容和淹没损失的自动估算等；通过与各专业部门数据的匹配分析，还可进行遥感影像地形畸变的自动校正，以及进行土地利用规划的研究等等。

1.1　地形因子的自动提取

根据空间矢量的分析原理，建立数字高程模型每一格网点的标准矢量 $\vec{P}_{i,j}$(图5-1)：

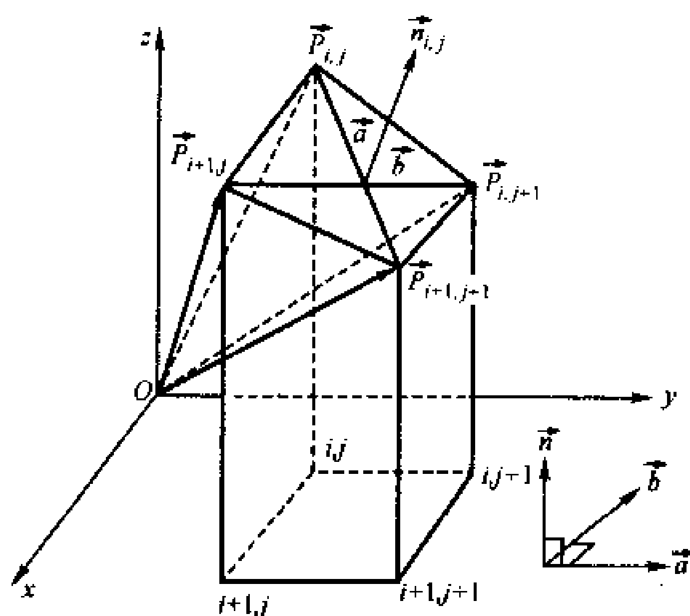


图5-1　空间矢量分析图

$$\vec{P}_{i,j}=\{(i-1)\times\Delta x+x_0,(j-1)\times\Delta y+y_0,z_{i,j}\}$$
$$i=1,2,\cdots,m-1,m;$$
$$j=1,2,\cdots,n-1,n。\qquad(5-1)$$

式中：Δx 为 x 轴方向的增量；

Δy 为 y 轴方向的增量；

x_0 和 y_0 为起始点坐标值；

$Z_{i,j}$为格网点高程值。

对于每个由相邻四个格网点确定的地表微分单元，其基本矢量 $\vec{a}$、$\vec{b}$ 计算公

式如下：

$$\vec{a}_{ij} = \vec{P}_{i+1,j+1} - \vec{P}_{i,j} = \{\Delta x, \Delta y, z_{i+1,j+1} - z_{i,j}\} \quad (5-2)$$

$$\vec{b}_{ij} = \vec{P}_{i,j+1} - \vec{P}_{i+1,j} = \{-\Delta x, \Delta y, z_{i,j+1} - z_{i+1,j}\} \quad (5-3)$$

基本矢量 $\vec{a}$、$\vec{b}$ 完全确定了微分单元在空间的特征。由 $\vec{a}$、$\vec{b}$ 可得地表单元法矢量 $\vec{n}_{ij}$：

$$\vec{n}_{ij} = \vec{a} \times \vec{b} = \begin{vmatrix} \vec{i} & \vec{j} & \vec{k} \\ x_a & y_a & z_a \\ x_b & y_b & z_b \end{vmatrix} = \begin{vmatrix} y_a & z_a \\ y_b & z_b \end{vmatrix} \vec{i} + \begin{vmatrix} z_a & x_a \\ z_b & x_b \end{vmatrix} \vec{j} + \begin{vmatrix} x_a & y_a \\ x_b & y_b \end{vmatrix} \vec{k}$$

$$= (y_a z_b - y_b z_a) i + (x_b z_a - z_b x_a) j + (x_a y_b - x_b y_a) k \quad (5-4)$$

将法矢量 $\vec{n}_{i,j}$ 由分矢量表示法转换为坐标表示法，则

$$\vec{n}_{i,j} = \{\Delta y(z_{i,j+1} + z_{i,j} - z_{i+1,j+1} - z_{i+1,j}), -\Delta x(z_{i,j+1} + z_{i+1,j+1} - z_{i+1,j} - z_{i,j}), +2\Delta x \Delta y\} \quad (5-5)$$

$$i = 1,2,3,\cdots,m-1; j = 1,2,3,\cdots,n-1。$$

同理，由空间微分单元边的中点所确定的矢量 $\vec{a}'$、$\vec{b}'$，可求得法矢量 $\vec{n}'_{i,j}$。根据法矢量，便可以进行地表单元各种地形因子的自动计算和提取。

1. 坡度计算

如图 5-2 所示，地表单元坡度就是其法矢量 $\vec{n}_{i,j}$ 与 z 轴之夹角，而两矢量夹角的余弦等于两矢量的数量积与模的乘积之商，即

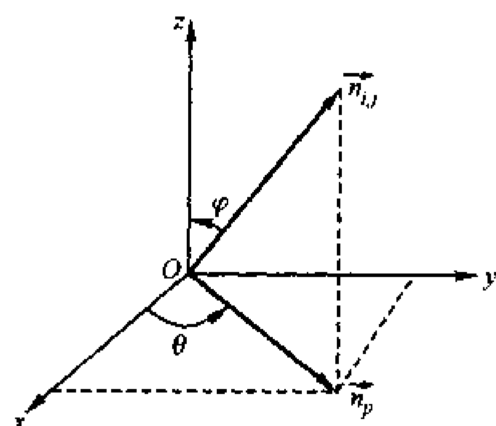


图 5-2　坡度和坡向的表示

$$\cos\varphi = \frac{\vec{z} \cdot \vec{n}_{i,j}}{|\vec{z}| \cdot |\vec{n}_{i,j}|} \quad (5-6)$$

设　$\vec{n}_{i,j} = \{x_1, y_1, z_1\}$

$\vec{z} = \{x_2, y_2, z_2\}$

则　$\vec{n}_{i,j} \cdot \vec{z} = x_1 x_2 + y_1 y_2 + z_1 z_2 = z_1$

$|\vec{n}_{i,j}| = \sqrt{x_1^2 + y_1^2 + z_1^2}$

$|\vec{z}| = \sqrt{x_2^2 + y_2^2 + z_2^2} = 1$

代入式(5-6)则有

$$\cos\varphi = \frac{z_1}{\sqrt{x_1^2 + y_1^2 + z_1^2}} =$$

$$\frac{2\Delta x \cdot \Delta y}{\{[\Delta y(z_{i,j+1} + z_{i,j} - z_{i+1,j+1} - z_{i+1,j})]^2 + [\Delta x(z_{i,j} + z_{i+1,j} - z_{i+1,j+1} - z_{i,j+1})]^2 + [2\Delta x \Delta y]^2\}^{\frac{1}{2}}}$$

则

$$\varphi = \arccos\{2\Delta x \cdot \Delta y / \{[\Delta y(z_{i,j+1} + z_{i,j} - z_{i+1,j+1} - z_{i+1,j})]^2 + [\Delta x(z_{i,j} + z_{i+1,j} - z_{i+1,j+1} - z_{i,j+1})]^2 + 4\Delta x^2 \cdot \Delta y^2\}^{\frac{1}{2}}\} \quad (5-7)$$

由坡度的概念知 $0° \leqslant \varphi \leqslant 90°$，故由式(5-7)容易确定坡度值。

当需要时，可将度数化为百分比表示，如图5-3所示。

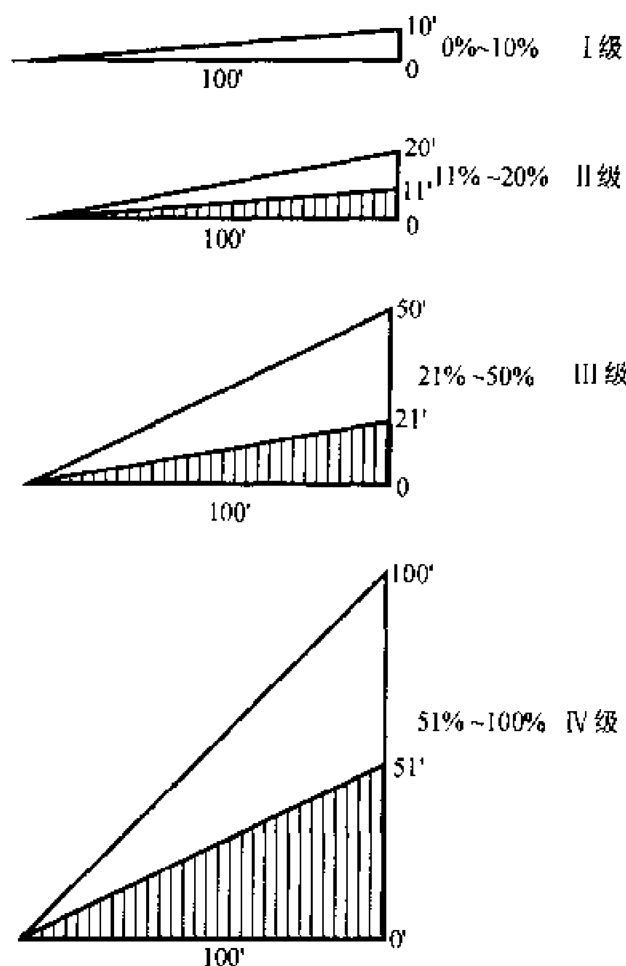


图5-3 坡度分级表示法

2. 坡向分析

在图5-2中，x 轴的方向为南，因此，地表单元坡向 θ 即为其法矢量 $\vec{n}_{i,j}$ 在 oxy 平面上的投影 $\vec{n}$ 与 x 轴的夹角。计算公式为

$$\begin{aligned}\theta &= \tan^{-1}(y_{\vec{n}_{i,j}} / x_{\vec{n}_{i,j}}) \\ &= \tan^{-1}(\Delta x * A(j) / \Delta y * B(j))\end{aligned} \quad (5-8)$$

式中：$A(j) = z_{(1,j+1)} + z_{(2,j+1)} - z_{(1,j)} - z_{(2,j)}$；

$B(j) = z_{(2,j)} + z_{(2,j+1)} - z_{(1,j)} - z_{(1,j+1)}$。

z 的下标变量如图5-4所示。

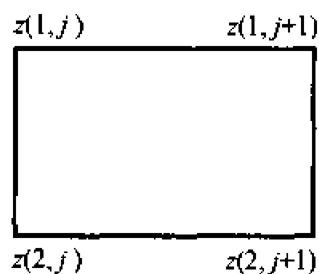


图5-4 坡向变量示意图

在实际应用中，由于所建立的 DEM 数据常常是按从南到北获取的（图 5-5），故在采用上述公式时，求出的 θ 是与正北方向的夹角。

另外，求出的坡向 θ 有与 x 轴正向与 x 轴负向夹角之分，这就要看坡向变量 $A(i)$和 $B(j)$的符号。在图 5-6 中，可以根据 θ 角及 $A(j)$和 $B(j)$的符号来确定地表单元的坡向，如表 5-1 所示。表中“≈”表示当 A 或 B 的绝对值很小时的情况，这与计算时的数值精度有关，可以根据情况设定一个阈值（例如 ε），当 $|A|$ 或 $|B|<\varepsilon$ 时，就可以认为 $|A|$ 或 $|B|=0$。根据该表求取的坡向类别比较详细，实际应用中有时反而不太方便，需要给予适当合并，例如分为平缓坡、阳坡、半阳坡和阴坡，并分别用代码 1、2、3 和 4 表示。

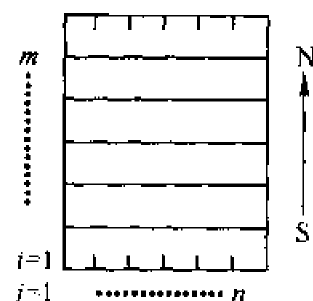


图 5-5 DEM 数据的获取方向

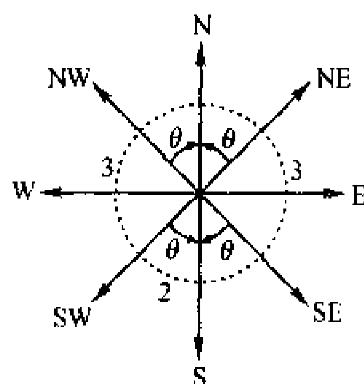


图 5-6 坡向的综合

表 5-1 坡向综合表

A	B	θ	坡 向	坡向合并	代 码
≈0	≈0		平缓坡	平缓坡	1
≈	<0	0	S	阳 坡	2
>0	<0	[-π/2,0]	WS		
<0	<0	[0,π/2]	ES		
>0	≈0	3/2π	W	半阳坡	3
<0	≈0	π/2	E		
≈0	>0	π	N	阴 坡	4
>0	>0	[π,3/2π]	WN		
<0	>0	[π/2,π]	EN		

3．曲面面积计算

地表单元的曲面面积很容易根据数字高程模型来取得，根据空间矢量的物理性质，单元曲面面积（$S_{i,j}$）可以用该单元边的中点所建立的矢量 $\vec{a}'_{i,j}$、$\vec{b}'_{i,j}$ 及由它们所确定的法矢量 $\vec{n}'_{i,j}$ 的模来定义，即

$$
\begin{aligned}
S_{i,j} &= |\vec{n}'_{i,j}| \\
&= [\{\Delta y^2(z_{i,j}+z_{i,j+1}-z_{i+1,j}-z_{i+1,j+1})^2\}/4 \\
&\quad +\{\Delta x^2(z_{i,j+1}+z_{i+1,j+1}-z_{i,j}-z_{i+1,j})^2\}/4 \\
&\quad +\Delta x^2\Delta y^2]^{\frac{1}{2}}
\end{aligned}
\tag{5-9}
$$

利用该公式计算了一个县的曲面面积为 2 576.418km²，而该县的投影面积为 2 419.313 km²，它是曲面面积的 94%，说明该县的地形属于起伏不大的丘陵区。

4．地表粗糙度计算

地表粗糙度是反映地表的起伏变化与侵蚀程度的指标，一般定义为地表单元的曲面面积与投影面积之比，这样，对于实际光滑的斜面也可以求出不同的粗糙度，这不适宜。这里用对顶点连线 L_1 与 L_2 中点的高差 D 来表示粗糙度（图 5－7），D 愈大，说明单元的四个顶点的起伏变化也愈大。其计算公式为

$$
\begin{aligned}
R_{i,j} = D &= |(z_{i+1,j+1}+z_{i,j})/2-(z_{i,j+1}+z_{i+1,j})/2| \\
&= \frac{1}{2}|z_{i+1,j+1}+z_{i,j}-z_{i,j+1}-z_{i+1,j}|
\end{aligned}
\tag{5-10}
$$

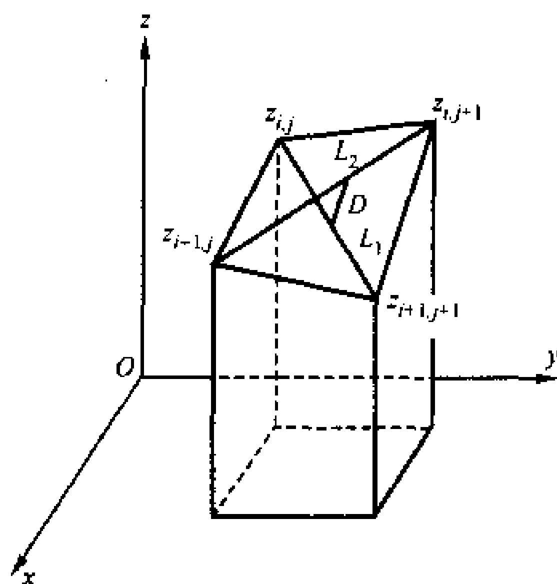


图 5－7 地表粗糙度计算

5．高程及变异分析

高程分析包括平均高程和相对高程的计算。以地表单元网格顶点 p_k（$k=1,2,3,4$）的高程平均值定义为该单元的平均高程。

$$
\bar{z} = \frac{1}{4}\sum_{k=1}^{4} z(p_k) \tag{5-11}
$$

以地表单元网格顶点 p_k（$k=1,2,3,4$）的高程与研究区域或某一流域内最

低点高程 z_{min} 之差的平均值定义为该单元的相对高程。

$$D_z = \frac{1}{4}\sum_{k=1}^{4} \quad z(p_k) - z_{min} \qquad (5-12)$$

高程变异是反映地表单元格网各顶点高程变化的指标，它以格网单元顶点的标准差与平均高程的比值来表示。

$$V = s/\bar{z} \qquad (5-13)$$

式中：标准差 $s = \left[\frac{1}{4}\sum_{k=1}^{4}(z(p_k) - \bar{z})^2\right]^{\frac{1}{2}}$。

6. 谷脊特征分析

谷和脊是地表形态结构的主要组成部分，由沟谷和山脊构成的地性线在区域地形的研究和制图综合中，都具有重要的意义。从一般概念上讲，谷即为地势相对最低的点集，脊为地势相对最高的点集。当对谷脊特征作概略分析时，可根据数字高程模型，按照以下判别式直接提取谷点和脊点。

如果 $(z_{i,j-1} - z_{i,j})(z_{i,j+1} - z_{i,j}) > 0$

当 $z_{i,j+1} > z_{i,j}$ 则 $VR(i,j) = -1$ (5-14)

当 $z_{i,j+1} < z_{i,j}$ 则 $VR(i,j) = 1$ (5-15)

如果 $(z_{i-1,j} - z_{i,j})(z_{i+1,j} - z_{i,j}) > 0$

当 $z_{i+1,j} > z_{i,j}$ 则 $VR(i,j) = -1$ (5-16)

当 $z_{i+1,j} < z_{i,j}$ 则 $VR(i,j) = 1$ (5-17)

如果(5-14)和(5-17)或(5-15)和(5-16)同时成立 则 $VR(i,j) = 2$

如果以上条件均不成立 则 $VR(i,j) = 0$ (5-18)

式中：
$$VR(i,j) = \begin{cases} -1, & \text{表示谷点} \\ 1, & \text{表示脊点} \\ 2, & \text{表示鞍点} \\ 0, & \text{表示其他点} \end{cases}$$

当需对谷脊特征作较精确的分析时，通过曲面拟合方法建立地表单元的曲面方程，然后通过确定曲面上各个插值点的极小值和极大值，以及当插值点在两个相互垂直的方向上分别为极大值和极小值时，可以确定出谷点、脊点或鞍点。

提取出谷点和脊点，将地表单元内所有谷点在单元域内的延伸长度累加，便获得单元的沟谷总长度 $\sum L$。沟谷总长度与地表单元面积 A 之比，即为沟谷密度。沟谷密度的精度随着与提取谷脊信息时给定的插值间距的缩小而提高。地表单元内几个谷点的切割深度的均值即为该单元的沟谷深度值，它以地表单元谷点与最近脊点的平均高差来确定，即

$$VH = \sum_{i=1}^{n}(\bar{h}_{R_i} - \bar{h}_{v_i})/n \qquad (5-19)$$

式中：h_{v_i} 为谷点高程；

h_{R_i} 为距该谷点最近的脊点的平均高程值。

7. 日照强度的分析

日照强度的分析是指在一个已知的地理区域内，通过坡度、方位和太阳在不同季节以及一天内的不同时间所处的位置等参数，来计算研究区每一点在某一时刻接受的日照强度。其计算公式为

$$E = \begin{cases} \beta \cdot G \cdot \sin h(a \cdot \cos t + b \cdot \sin t + \cos\theta \cdot \sin h) \\ 0 \end{cases} \tag{5-20}$$

当 $E = 0$ 时，表示阴暗；当 $0 < E < 100$ 时表示不同的日照强度。

式中：β 是大气透过率（与太阳高度和大气状况有关）；

G 为太阳常数；

h 为太阳高度角（由球面三角公式推求）；

t 是时角；

a 和 b 为坡面方程系数；

θ 为坡度。

8. 淹没边界的计算

DEM 数据不仅直接用于各种地形因子的自动提取，而且可以与有关信息进行复合，研究地形因子与其他要素之间的相关关系，其中淹没损失的估算，就是研究土地利用数据与数字高程之间的关系。为了估算淹没损失，首先，将数字地形的数据和土地利用数据进行匹配；其次，根据高程，确定淹没边界；最后，统计淹没边界内的被淹对象的类别和面积，就能精确地算出研究对象的淹没损失。为确定淹没边界，如图 5-8，设格网边长为 Δx，洪水的实际淹没高程为 H，该高程附近的平均坡度为 a，于是

$$H_1 = H - \frac{1}{2}\Delta x \mathrm{tg} a$$

和

$$H_2 = H + \frac{1}{2}\Delta x \mathrm{tg} a$$

显然，由高程 H_1 确定的网格正好在淹没边界之内，而由 H_2 确定的网格，则正好超过实际淹没高程 H 一个网格。由于在数字地形模型中，相邻网格的高程总是存在于 H_1 至 H_2 的高程范围内，因此，凡高程 $< H_2$ 的格网均可作为淹没区处理，只是对于边界网格，可能存在半个网格的误差，需要进行误差改正。以 H_2 为临界值，将数字地形的数据按照下式进行重新分类，即

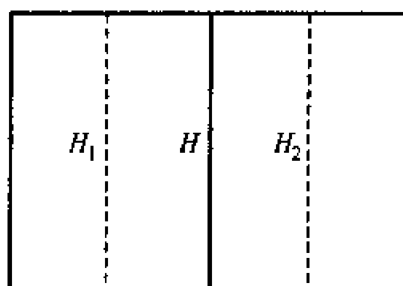


图 5-8 淹没边界的计算

$$z_{i,j} = \begin{cases} 1 & 当\ z_{i,j} \leqslant H_2 \\ 0 & 当\ z_{i,j} > H_2 \end{cases}$$

得到数字地形的二值图，将该二值图与土地利用数据进行布尔逻辑操作，就能统计出各类土地利用的淹没面积。

1.2 地表形态的自动分类

在 DEM 数据文件的基础上进行地形的自动分类，首先根据区域的地形特点，拟定地形分类决策表（表 5－2），然后按照自动提取地形类型信息的过程（图 5－9），便可获得区域的地形类型分类系统（例如岗丘、丘陵、丘陵阳坡地、丘陵阴坡地等），并输出地形类型图（图 5－10）。

表 5－2 地形分类决策表

	平 地	岗 丘	丘 陵	低 山	中 山
绝对高度 H/m			<400	400～800	>800
相对高度 ΔH/m		<100	100～200	>200	>200
坡度 s	<3°				

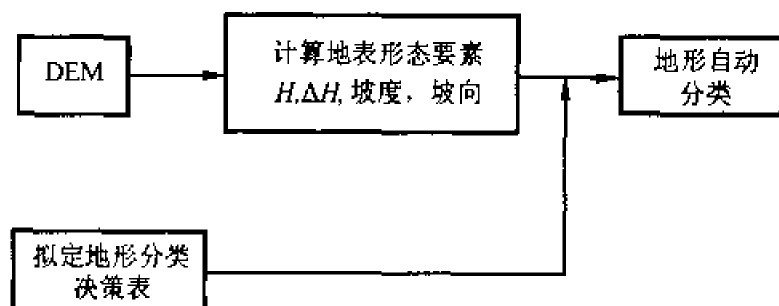


图 5－9 自动提取地形类型信息过程

1.3 地学剖面的绘制和分析

研究地学剖面，常常可以以线代面，概括研究区域的地势、地质和水文特征，包括区域内的地貌形态、轮廓形状、绝对与相对高度、地质构造、斜坡特征、地表切割强度和侵蚀因素等等。如果在地形剖面上叠加表示其他地理变量，例如坡度、土壤、岩石抗蚀性、植被覆盖类型、土地利用现状等，可以作为提供大地侵蚀速率研究、农业生产布局的立体背景分析、土地利用规划、工程决策（例如工程选线和位址选择）等的参考依据。这里提供自动绘制地形剖面图和各种类型综合剖面图的方法，包括任意地形剖面的内插算法，剖面上地理变量的自动叠加和表示，以及剖面图的解释和应用。地理剖面图的自动绘制和表示，是区域栅格数据

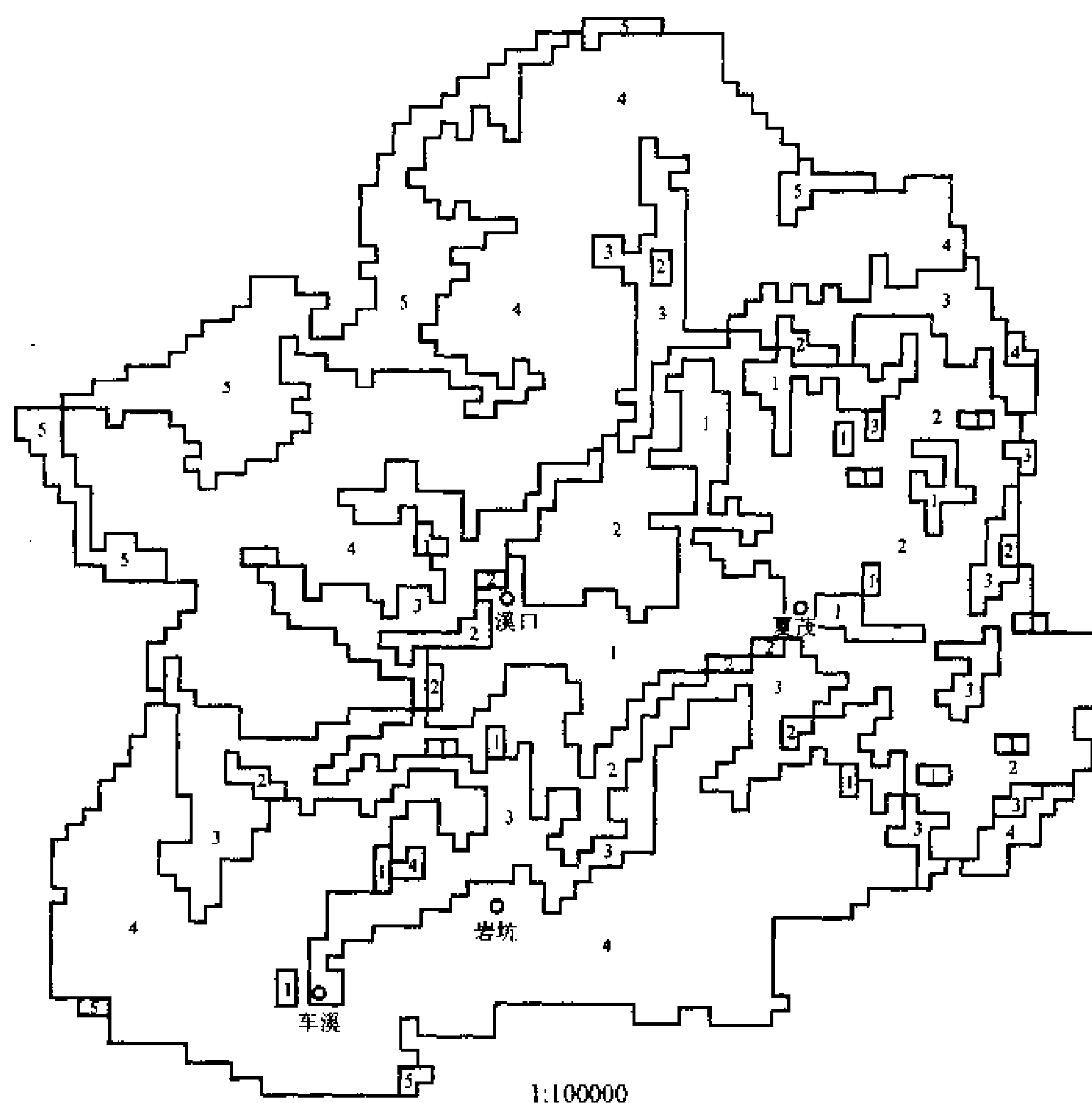


图5-10　地形分类输出

1. 平地　　2. 岗丘　　3. 丘陵
4. 低山　　5. 中山

的应用内容之一，是地理信息系统分析工具的组成部分，它为区域地理数据的处理和立地条件的分析，提供了一种有效的研究手段。

如图5-11所示，设 $z_{i,j}$ 为网格点 (i,j) 上的海拔高度，它可以根据研究地区的地形图，通过数字化仪沿等高线采样，取得离散数据 $D_k(x_k, y_k)$，由 D_k 的 z 值通过按距离加权平均或分块函数法，自动建立研究地区的数字高程模型DEM，并可写成矩阵形式

$$\mathrm{DEM} = [z_{i,j}], \quad i = 1,2,\cdots,m; \quad j = 1,2,\cdots,n$$

根据 $[z_{i,j}]$ 数据来绘制地形剖面图，只要知道所绘剖面线在数字高程模型中

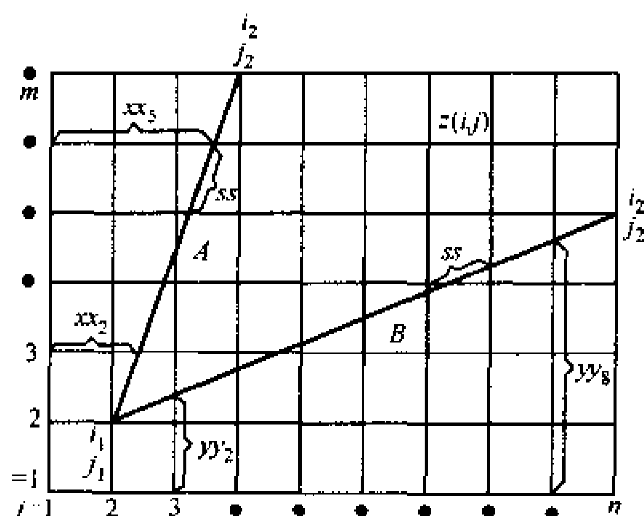


图 5-11 剖面线交点的内插

的起点位置(i_1,j_1)和终点位置(i_2,j_2),而且

$$i \leqslant i_1、i_2 \leqslant m$$

$$j \leqslant j_1、j_2 \leqslant n$$

就可以惟一地确定这条剖面线与 DEM 网格各个交点的平面位置及其高程(图 5-11),设 $\Delta x = j_2 - j_1, \Delta y = i_2 - i_1$,显然,当 $\Delta x \neq 0$,且 $|\Delta y/\Delta x| - 1 \geqslant 0$ 时,应求剖面线与 DEM 网格横轴的交点(图 5-11),它们在 DEM 坐标系中的位置和高程分别为

$$yy_k = i_1 + (k-1) \times \mathrm{ISG}_2 \tag{5-21}$$

$$xx_k = j_1 + |(yy_k - i_1)/(i_2 - i_1) \times (j_2 - j_1)| \times \mathrm{ISG}_1 \tag{5-22}$$

$$zz_k = (xx_k - IA)(z_{IK,IB} - z_{IK,IC}) + z_{IK,IC} \tag{5-23}$$

式中: $IK = [yy_k]$;

$IA = [xx_k]$;

$IB = (IA+1) \times \mathrm{ISG}_1$;

$IC = IB - \mathrm{ISG}_1$;

| |表示取绝对值;

[ ]表示取整数值;

$k = 2, 3, \cdots, |i_2 - i_1|$。

ISG_1 和 ISG_2 的值由 Δx 和 Δy 的符号来决定(表 5-3),它表示剖面线起点和终点的位置,可以按照任意的方向来设置。

同理,当 $\Delta x \neq 0$,且 $|\Delta y/\Delta x| - 1 < 0$ 时,应求剖面线与 DEM 网格纵轴的交点(图 5-11),它们在 DEM 坐标系中的位置和高程分别为

表 5-3 ISG_1 和 ISG_2 值的确定

Δx	>0	>0	<0	<0
Δy	>0	<0	>0	<0
ISG_1	1	1	-1	-1
ISG_2	1	-1	1	-1

$$xx_k = j_1 + (k-1) \times ISG_1 \tag{5-24}$$

$$yy_k = i_1 + |(xx_k - j_1)/(j_2 - j_1) \times (i_2 - i_1)| \times ISG_2 \tag{5-25}$$

$$zz_k = (yy_k - IA)(z_{IB,IK} - z_{IC,IK}) + z_{IC,IK} \tag{5-26}$$

式中：$IK = [xx_k]$；

$IA = [yy_k]$；

$IB = (IA + 1) \times ISG_2$；

$IC = IB - ISG_2$；

$k = 2, 3, \cdots, |j_2 - j_1|$。

当 $\Delta x = 0$ 时，表示剖面线方向与DEM网格纵轴方向相一致，因此剖面线上各点的高程可以直接写出

$$zz_k = z_{IB,IC} \tag{5-27}$$

式中：$IB = i_1 + (k-1) * ISG_2$；

$IC = i_1$ 或 j_2；

$k = 1, 2, \cdots, |i_2 - i_1| + 1$。

同理，当 $\Delta y = 0$ 时，表示剖面线方向与DEM网格的横轴相一致，而且剖面线上各点的高程为

$$zz_k = z_{IB,IC} \tag{5-28}$$

式中：$IB = i_1$ 或 i_2；

$IC = j_1 + (k-1) \times ISG_1$；

$k = 1, 2, \cdots, |j_2 - j_1| + 1$。

计算出剖面线上各点的高程 zz_k(m)和剖面线相邻两点的实际距离 ss(m)，就可以根据选定的垂直比例尺 m_y 和水平比例尺 m_x，自动绘出所需要的地形剖面图(图5-12)，剖面线相邻两点的实际距离 ss，可以根据以下公式给出

$$SS = \begin{cases} \sqrt{(\Delta x/\Delta y \times sx)^2 + sy^2} & \text{当 } \Delta x \neq 0, \left|\frac{\Delta y}{\Delta x}\right| - 1 > 0 \\ \sqrt{(\Delta y/\Delta x \times sy)^2 + sx^2} & \text{当 } \Delta x \neq 0, \left|\frac{\Delta y}{\Delta x}\right| - 1 < 0 \\ \sqrt{sx^2 + sy^2} & \text{当 } \left|\frac{\Delta y}{\Delta x}\right| = 1 \end{cases} \tag{5-29}$$

式中：sy(DEM网格单位纵边长，m)，当 $\Delta x = 0$ 时；

sx(DEM网格单位横边长，m)，当 $\Delta y = 0$ 时。

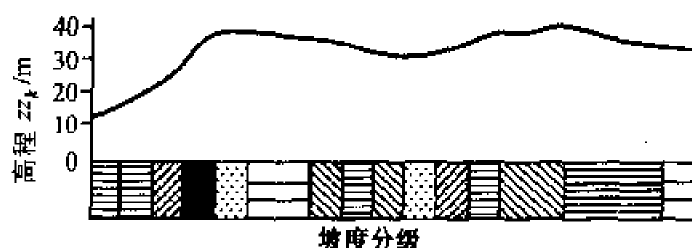


图 5-12 沿某一方向绘出的典型剖面图

§2 空间叠合分析

2.1 空间叠合分析的概念

空间叠合分析(spatial overlay analysis)是指在统一空间参照系统条件下,每次将同一地区两个地理对象的图层进行叠合,以产生空间区域的多重属性特征,或建立地理对象之间的空间对应关系。前者一般用于搜索同时具有几种地理属性的分布区域,或对叠合后产生的多重属性进行新的分类,称为空间合成叠合(图 5-13);后者一般用于提取某个区域范围内某些专题内容的数量特征,称为空间统计叠合(图 5-14)。

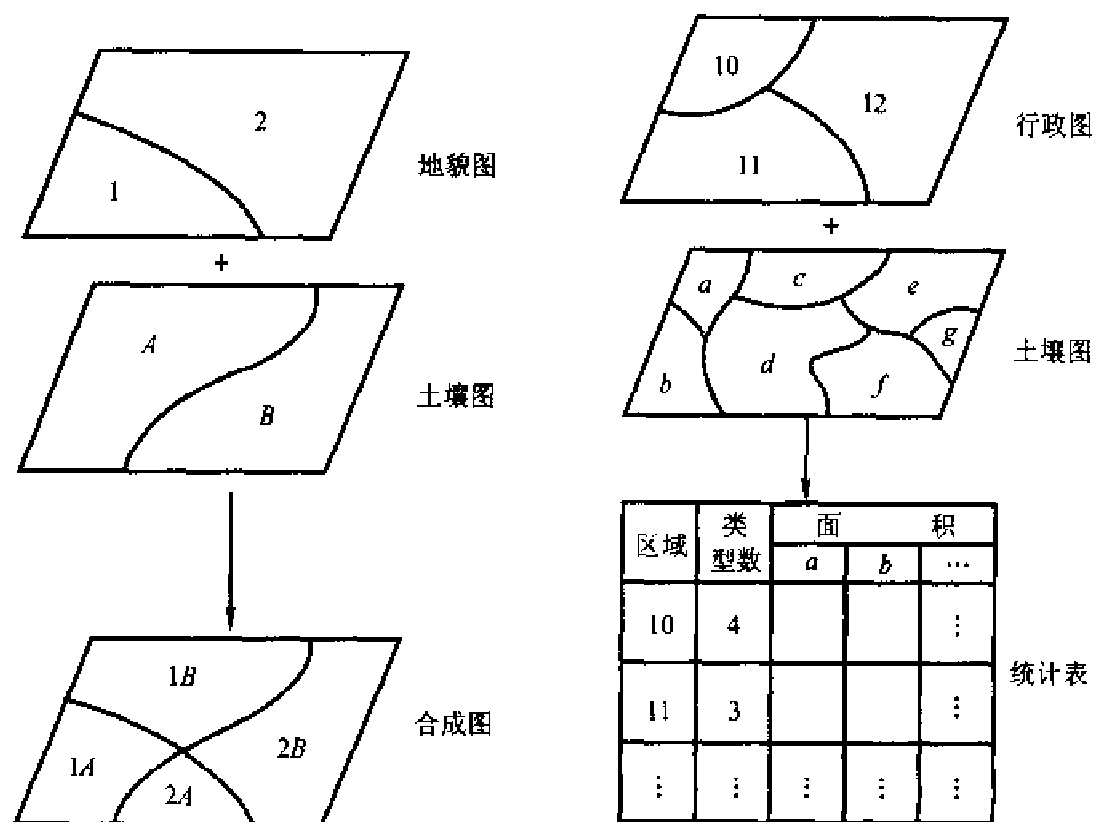


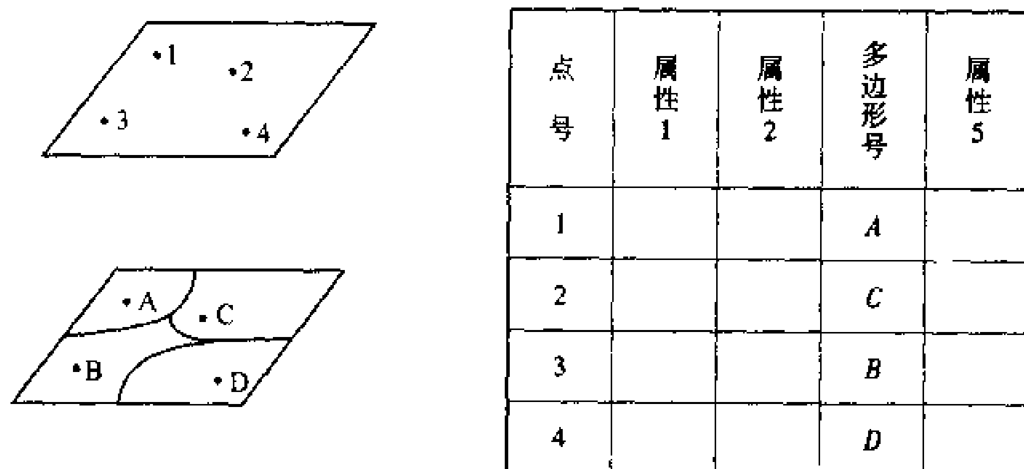
图 5-13 合成叠合

图 5-14 统计叠合

空间叠合分析根据叠合对象图形特征的不同，分为点与多边形的叠合(point - in - polygon overlay)、线与多边形的叠合(line - in - polygon overlay)和多边形与多边形的叠合(polygon - on - polygon overlay)三种类型。

1. 点与多边形的叠合

点与多边形的叠合是确定一图层上的点落在另一图层的哪个多边形内，以便为图层的每个点建立新的属性。例如水井点位与规划区多边形相叠合，可确定每口井所属的规划区范围，它实质是点与面之间的包含分析，可采用铅垂线算法来实现点与多边形的叠合分析(图5-15)。

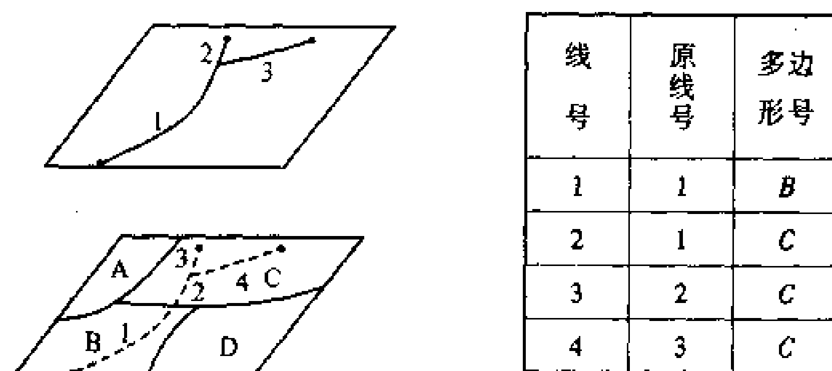


点号	属性1	属性2	多边形号	属性5
1			*A*	
2			*C*	
3			*B*	
4			*D*	

图5-15 点与多边形叠合分析

2. 线与多边形的叠合

线与多边形的叠合是确定一图层上的弧段落在另一图层的哪个多边形内，以便为图层的每条弧段建立新的属性。例如当确定某一行政区内各种等级道路的里程数时，就需要将道路图与境界图相叠合，计算弧段与多边形边界的交点，在交点处截断弧段，并对弧段重新编号，建立弧段与多边形的归属关系，如图5-16所示。



线号	原线号	多边形号
1	1	*B*
2	1	*C*
3	2	*C*
4	3	*C*

图5-16 线与多边形叠合分析

3. 多边形与多边形的叠合

多边形与多边形的叠合是指将两个不同图层的多边形要素相叠合，产生输

出层的新多边形要素,用以解决地理变量的多准则分析、区域多重属性的模拟分析、地理特征的动态变化分析,以及图幅要素更新、相邻图幅拼接、区域信息提取等。因此,多边形与多边形的叠合分析具有广泛的应用功能,它是空间叠合分析的主要类型,一般基础 GIS 软件都提供该类型的叠合分析功能,以 ARC/INFO 为例,提供的多边形与多边形叠合分析功能,包括以下六种操作命令:

(1) Union。进行多边形叠合,输出层为保留原来两个输入图层的所有多边形。显然,如果两个输入的图层分别代表两个不同时期的河道要素,则输出的图层保留了该河道经历了两个变迁时期形成的所有特征(图 5-17)。

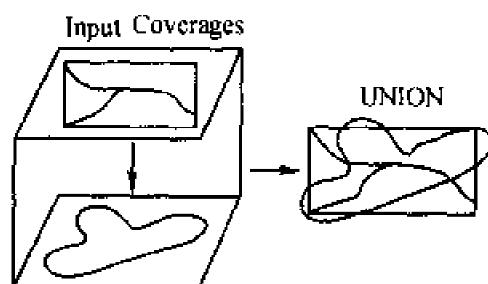


图 5-17 Union 叠合操作

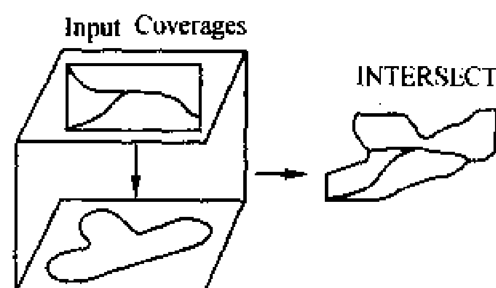


图 5-18 Intersect 叠合操作

(2) Intersect。进行多边形叠合,输出层为保留原来两个输入图层的共同多边形。显然,这时上述两个图层叠合后,输出的结果为表示该河道经历两个变迁时期都不发生变化的部分(图 5-18)。

(3) Identity。进行多边形叠合,输出层为保留以其中一输入图层为控制边界之内的所有多边形。显然,这时两个原始图层叠合后,输出的结果能清晰地反映出该河道经过两个时期以后动态变化的特征(图 5-19)。

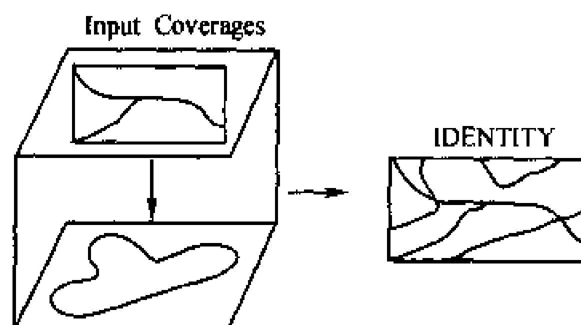


图 5-19 Identity 叠合操作

(4) Erase。进行多边形叠合,输出层为保留以其中一输入图层为控制边界之外的所有多边形。显然,这时表示在将更新的特征加入之前,须将控制边界之内的内容删除(图 5-20)。

(5) Update。进行多边形叠合,输出层为一个经删除处理后的图层与一个新

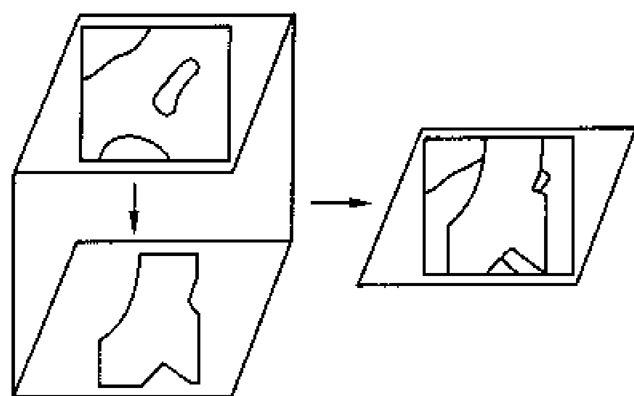


图 5-20 Erase叠合操作

特征图层进行合并后的结果(图 5-21)。

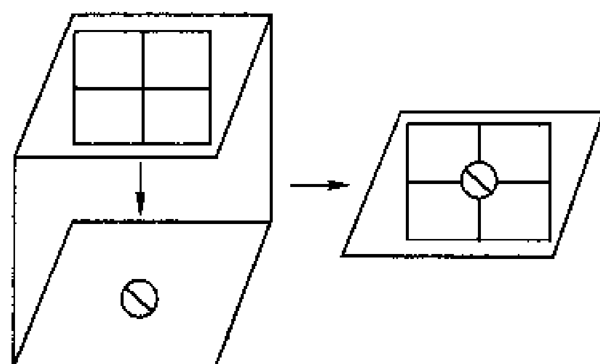


图 5-21 Update叠合操作

(6) Clip。进行多边形叠合,输出层为按一个图层的边界,对另一个图层的内容要素进行截取后的结果(图 5-22)。

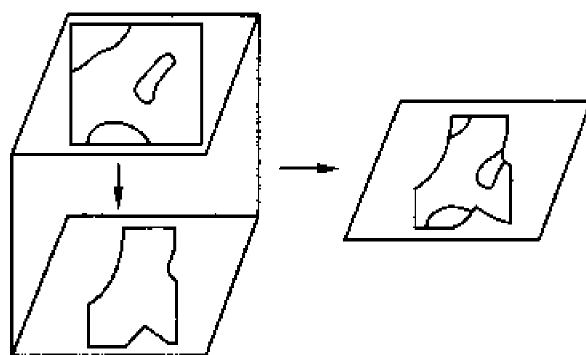


图 5-22 Clip叠合操作

空间叠合分析根据所采用的数据结构的不同,分为基于矢量数据的叠合分析和基于栅格数据的叠合分析两种类型。

2.2 基于矢量数据的叠合分析

基于矢量数据的叠合分析是参与分析的两个图层的要素均为矢量数据。矢量数据的叠合算法，虽然数据存储量比较小，但是运算过程比较复杂。设参与叠合的两个图层均为多边形，其中被叠合的多边形为本底多边形，用来叠合的多边形为上覆多边形，叠合后产生具有多重属性的多边形，称为新多边形，则基于矢量数据的多边形叠合算法过程如图 5－23 所示：

2.3 基于栅格数据的叠合分析

基于栅格数据的叠合分析是参与分析的两个图层的要素均为栅格数据。栅格数据的叠合算法，虽然数据存储量比较大，但是运算过程比较简单。如图 5－24所示，设参与叠合分析的两个图层的变量为 U 和 V，则建立对应于 U 和 V 的某行（箭头所示）的游程编码数据（表 5－4）可以表示为：

表 5－4 游程编码数据

变量	游程编码数据			
U	A_i	0	P_i	329
		1		336
		2		345
V	A_j	0	P_j	329
		2		330
		1		338
		2		345

$$U = (A_i, P_i), i = 1, m$$

$$V = (A_j, P_j), j = 1, n$$

式中：A_i、A_j 分别为变量 U 和 V 某行各段游程的属性值；

P_i、P_j 分别为变量 U 和 V 该行各段游程最后一列的列号；

m、n 分别为该行游程的总数。

于是，按照图 5－25 所示的算法，可以迅速实现变量 U 和 V 的叠合分析。

由叠合分析建立的数据文件（表 5－5），为形成满足用户某种应用需求的结果，需要根据原始变量，按照某种指标，建立属性转换矩阵，然后根据该属性转换矩阵对数据文件进行属性的重新分类。表 5－6 是按照某种车辆通行影响因子所建立的属性转换矩阵。图 5－26 是根据该属性转换矩阵的分类结果输出的专题地图。

通常，基于栅格数据的叠合分析根据应用目的的不同，可以利用以下三种变换方法对参与叠合的图层进行新属性值的计算。

1. 点变换方法

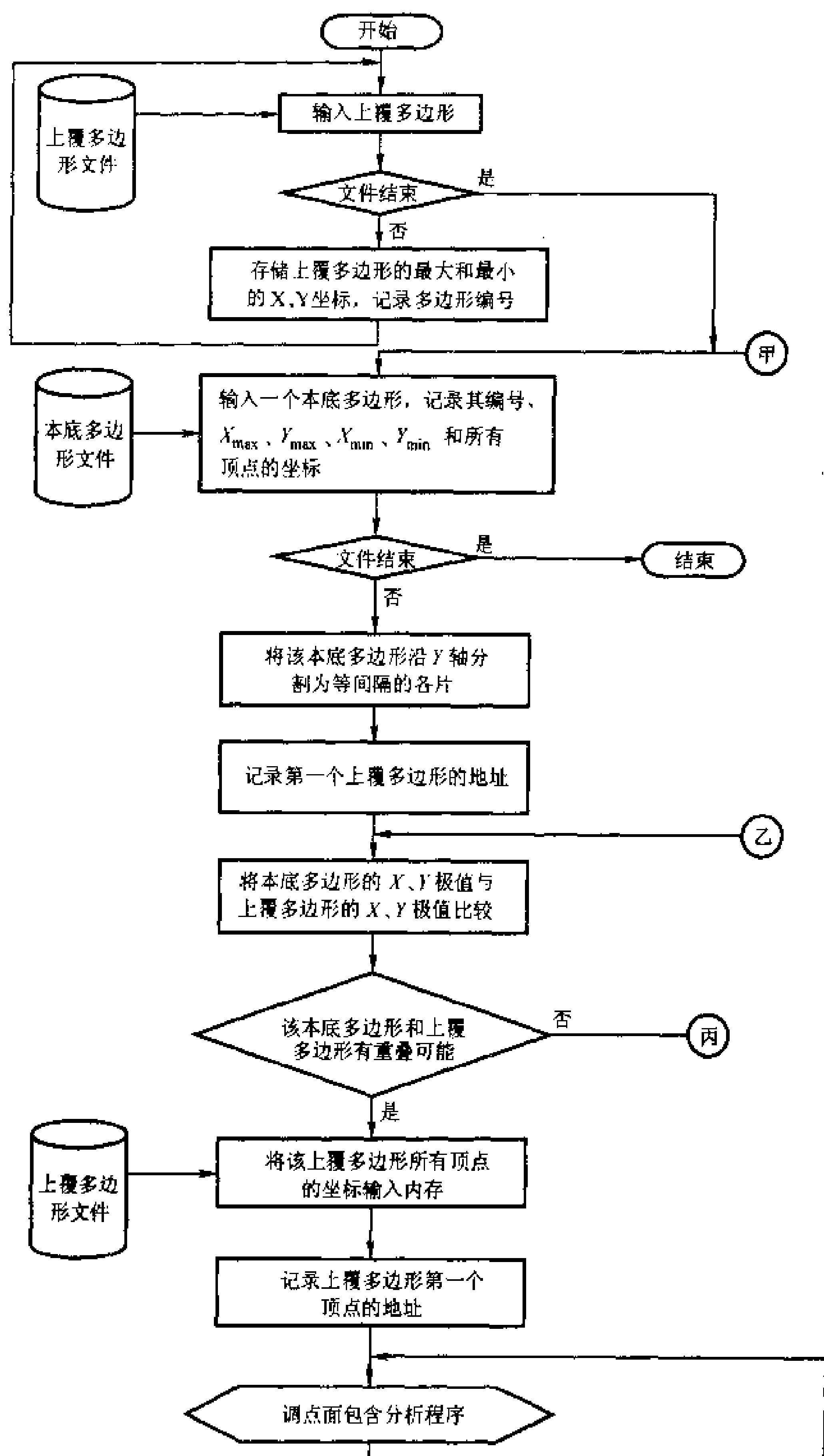
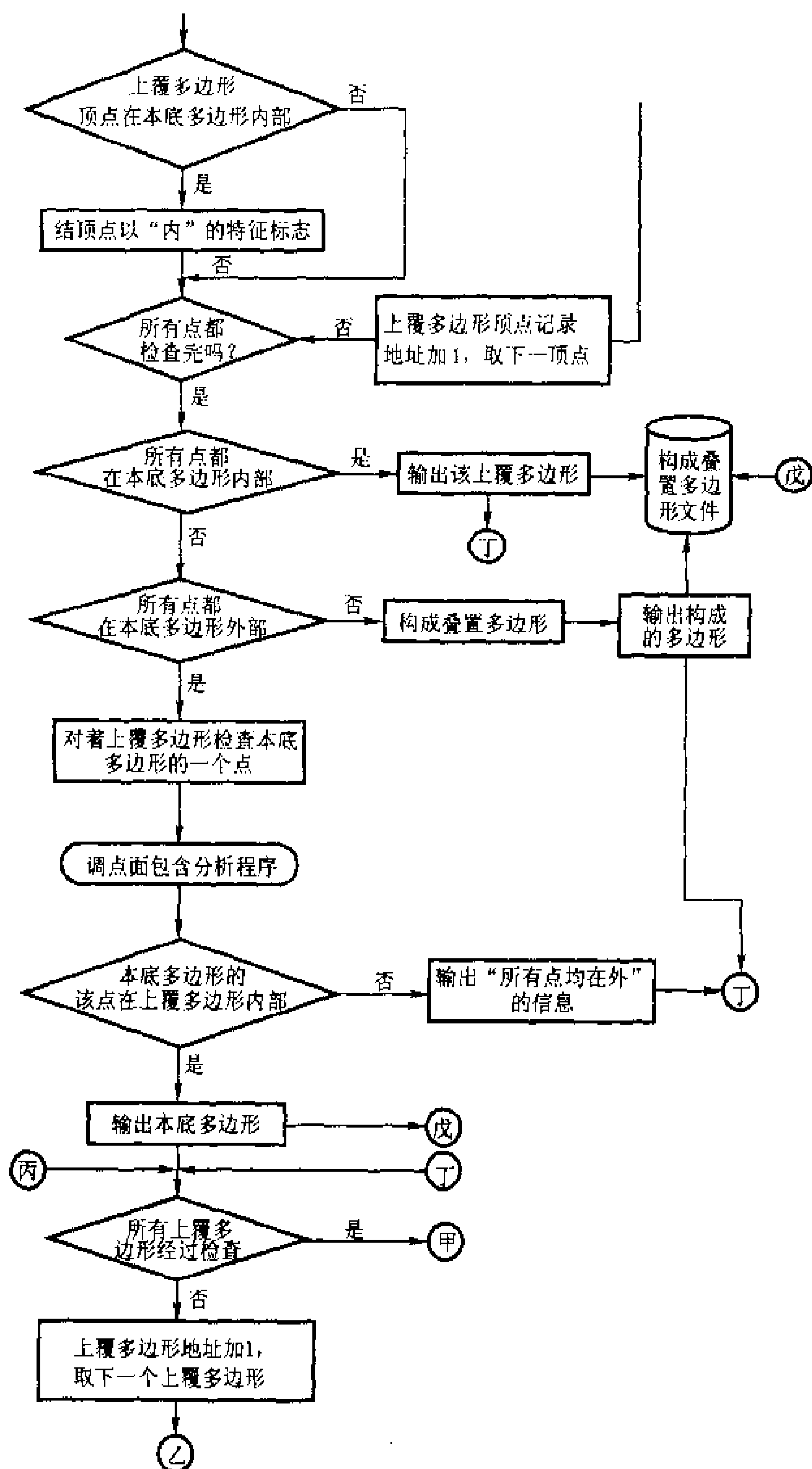
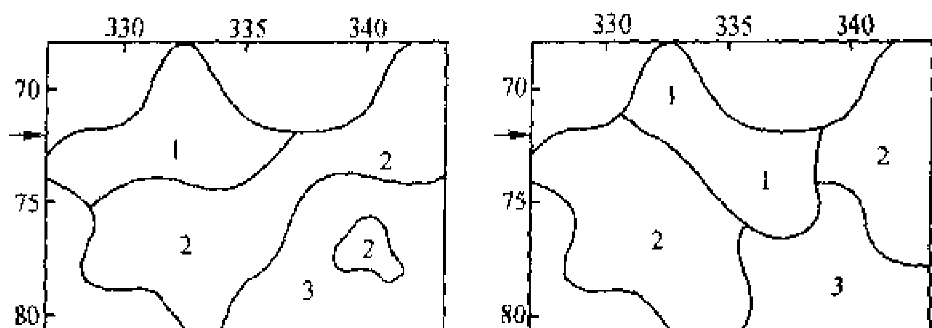


图5-23　基于矢量数据的多


上覆多边形
顶点在本底多边形内部
否
是
结顶点以“内”的特征标志
否
所有点都
检查完吗?
否
上覆多边形顶点记录
地址加1，取下一顶点
是
所有点都
在本底多边形内部
是
输出该上覆多边形
构成叠
置多边
形文件
戊
丁
否
所有点都
在本底多边形外部
否
构成叠置多边形
输出构成
的多边形
是
对着上覆多边形检查本底
多边形的一个点
调点面包含分析程序
本底多边形的
该点在上覆多边形内部
否
输出“所有点均在外”
的信息
丁
是
输出本底多边形
戊
丙
丁
所有上覆多
边形经过检查
是
甲
否
上覆多边形地址加1，
取下一个上覆多边形
乙

边形叠合算法流程图



(a)变量 U(坡度图)
1. 平坦　2. 中等　3. 陡峻

(b) 变量 V(土壤图)
1. 岩石　2. 砂土　3. 粘土

图 5-24　栅格叠合分析的原始变量

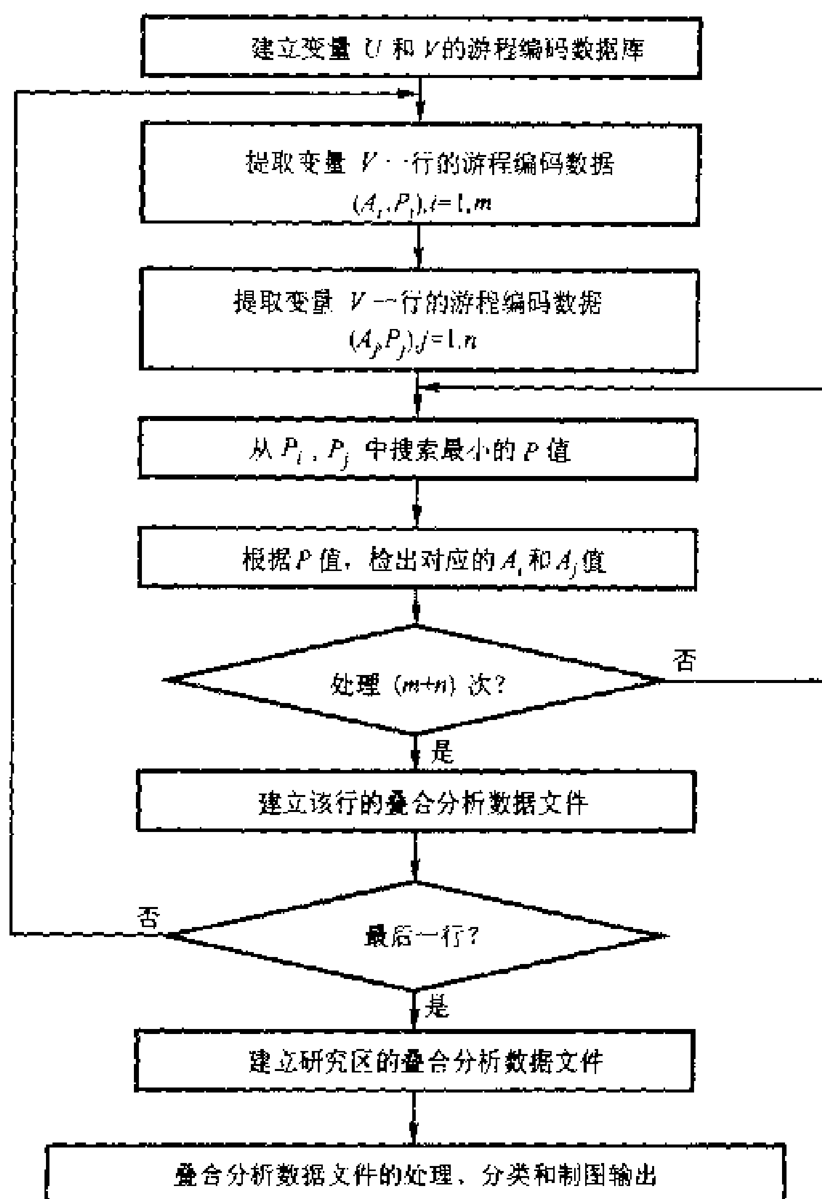


图 5-25　基于栅格数据的叠合分析

表 5-5 由叠合分析建立的数据文件

P	A_U	A_V	P	A_U	A_V
329	0	0	338	2	1
330	1	2	345	2	2
336	1	1	⋮	⋮	⋮

表 5-6 基于车辆通行能力的属性转换矩阵

变量 U \ 变量 V		土壤类型		
		1 岩石	2 砂土	3 粘土
坡度类型	1 平坦	容易	容易	容易
	2 中等	容易	困难	一般
	3 陡峻	一般	困难	困难

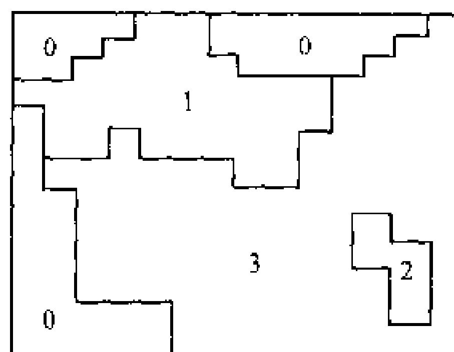


图 5-26 车辆通行能力分析图

0 界外区 1 容易 2 一般 3 困难

点变换只依据参与叠合图层相应点的属性值进行新属性值的运算，既与各图层邻域点的属性值无关，也不受区域内一般特征的影响。其运算函数可以是加、减、乘、除、指数、对数、三角函数等，也可以是由基本函数组成的复杂函数式等。运算得到的新属性值可能与原图层的属性意义完全不同，例如新变量或类型等。基于栅格叠合的点运算函数如表 5-7 所示。

表5-7　栅格叠合点运算函数

运算方法	关键字	修饰语	简　　述
加法	ADD	加权	各层上的属性相加得新层上相应点的值
减法	SUBTACT	加权	两层上对应点的属性相减
乘法	MULTIPLY	加权	相乘
除法	DIVIDE	加权	相除
乘方	EXPONENTIATE	加权	乘方
三角函数			变换成三角函数
逻辑函数			变换成逻辑函数
平方根			取平方根值
平均值	AVERAGE		取加权或不加权的平均值
覆盖	COVER		用第二幅地图上的非零像元覆盖第一幅地图产生第三幅
逻辑组合	CROSS		把多幅图的属性按布尔逻辑关系组合成新图上的属性
查找范围	MAXIM12E MINIM12E		找出多幅图上属性的最大或最小值作为第三图幅上的新值
提取或分离	EXTRACT		从一层上选取特定属性值或值的区间产生新层
确定频率分布	HISTOGRAM		计算直方图
频率分布与理论分布匹配	DISTRIBUTION STRETCH		重新计算像元值使其分布与理论分布一致
赋常数	CONSTANT		将一层上的所有像元赋予同一个常数
分　类	SLICE		将连续值(或离散值)按数值区间分成类别
两层的双向比较	CROSS	平均值 总和 极大值 极小值 中值 多(少)数 差异	建立属性的交叉表格,主要是对两层比较后,以点对点为基础的统计结果产生新地图
多变量运算和类别定位	PCOMP		n层像元值的基本元素变换,并写到输出层上去
	CLUSTER		完成聚群分析
	ALLOCATE		按分类标准把像元分配到输出地图上的某类别
傅立叶变换+滤波	FOURIER		计算二维傅立叶变换并用滤波器滤除高或低频成分

2. 区域变换方法

区域变换是指在计算新图层相应栅格的属性值时，不仅与原图层对应栅格的属性值有关，而且要顾及原图层栅格所在区域的几何特性(区域长度、面积、周长、形状等)或原图层同名栅格的个数。例如，当进行空间统计叠合分析时，需要应用区域变换方法。有关栅格叠合的区域变换函数如表5-8所示。

表5-8 栅格叠合区域变换函数

运算方法	关键字	简 述
重分类或重编码	RENUMBER	对指定的单元或图例中的值或地图上的数值范围赋予新值，此法也可用于图像反差增强
确定数量	REPORT SIZE	计算或从数据库中读取区域的相应数据并显示计算结果或写进文件
确定区域的形状	RULER	确定一定区域的形状和某一层上分离开的区域数量
按面积分类	SIZE	按计算出的每个区域的面积大小和输出地图的要求重新给予每个区域一个名称
两多边形网叠合	OVERLAY	两层多边形网络叠合产生新多边形网络(矢量或栅格均可)
按区域进行逻辑选择		与点运算相似
按区域的内容分类		与点运算相似

3. 邻域变换方法

邻域变换是指在计算新图层相应栅格的属性值时，不仅考虑原图层对应栅格及其属性值，而且还应顾及与该栅格相关联的邻域或影响半径范围内的栅格属性值的影响。例如，当进行空间离散点搜索、内插或缓冲带操作时，需要应用邻域变换方法。有关栅格叠合的邻域变换函数如表5-9所示。

表5-9 邻域变换函数及其应用

运算方法	关键字	修饰词	简 介
邻接单元的延续	CLUMP	距离限制	建立新地图时按惟一性识别标准确定延续的单元组，一般以距离作为识别标准
	SCAN	距离，均值 总和，极值 中间值、 偏差、 方差等	按地图上一定距离范围内邻域单元的统计特征计算每个单元的累积值，从而建立一幅新的地图
	SPAN	距离	按某像元到其他类别的像元之间的距离计算离散地图上每个像元的新值

续表

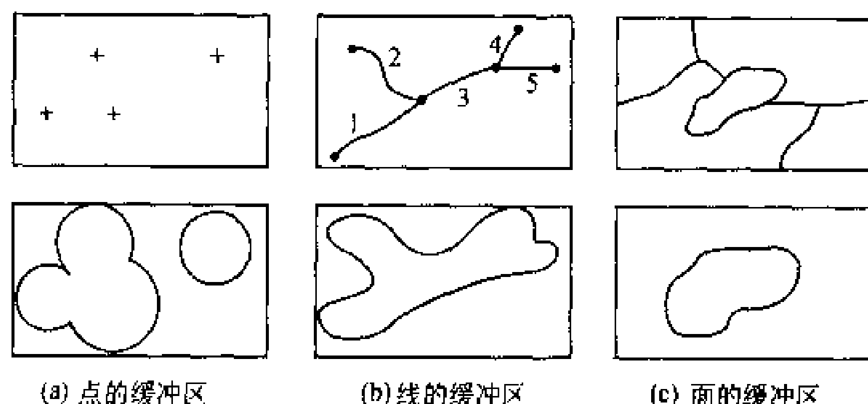
运算方法	关键字	修饰词	简　　介
找出给定目标的取值区间或距离内的位置	DSEARCH	距　离 区　间 方　向	在指定的距离和数值区间内，从连续或离散地图数据中选择出需要的点，建立新地图
连续表面函数	SLOPE ASPECT PROFILE	均　值 极大值	对表示连续表面的数据微分产生坡度图作为输出地图 同上，但主要是计算坡度方向 分析斜坡断面类型
由离散数据建立连续表面的函数	SPREAD	距　离 越　过 向　上 向下穿过	按到起始点或起始线的距离值重新编码产生连续地图 控制延伸方向，即用连续或者离散地图来建立延伸通道 基底地图也可能包括某些障碍物等
等值化	CONTOUR		从栅格数据建立等值线图
内　插	TREND INVERSE OPTIMAL		计算趋势面和地图(这是连续的叠置层) 计算每个像元的加权平均值，作为连续层，用协方差函数进行内插
搜寻最短通　道	VIEW STREAM PATHWAY		根据某层上的目标点向四周的连续三维表面观察能否看到的条件把其他像元(点)分成两类(能看到和不能看到) 搜寻连续表面上两点间的最短路线 按时间、费用、距离等在网络系统(如道路网络系统)中找出最短通道
点在多边形中	POINT		找出包含指定点的多边形

以上对基于栅格数据的叠合分析，讨论了三种主要的变换方法，在实际应用中可以通过交互运算，满足不同的空间分析需求。

§3 空间缓冲区分析

3.1 空间缓冲区分析的模型

空间缓冲区分析(spatial buffer analysis)是指根据分析对象的点、线、面实体,自动建立它们周围一定距离的带状区,用以识别这些实体或主体对邻近对象的辐射范围或影响度(图5-27),以便为某项分析或决策提供依据。它是地理信息系统重要的和基本的空间操作功能之一。例如,在城市研究中,当改变某个辖区的行政界线时,需要通知周围一定距离范围内(如150 m)的住户;在林业规划中,需要按照距河流一定纵深的范围来规划森林的砍伐区,以防止水土流失;在地震带,要按照断裂线的危险等级,绘出围绕每一断裂线的不同宽度的缓冲带,作为警戒线的指示;在土地评价时,要根据离开交通线的远近,进行成本的估算等等。这些都需要应用缓冲区的空间操作方法。



(a) 点的缓冲区　(b) 线的缓冲区　(c) 面的缓冲区

图5-27 缓冲区的建立
(据ESRI,1988)

在进行空间缓冲区分析时,通常要将研究的问题抽象为以下三类因素来进行分析:

(1) 主体。表示分析的主要目标,一般分为点源、线源和面源三种类型,如图5-27所示;

(2) 邻近对象。表示受主体影响的客体,例如行政界线变更时所涉及的居民区、森林遭砍伐时所影响的水土流失范围等等;

(3) 作用条件。表示主体对邻近对象施加作用的影响条件或强度。

根据主体对邻近对象作用性质的不同,一般可采用以下三种不同的分析模型:

① 线性模型。用于当主体对邻近对象的影响度(F_i)随距离(r_i)的增大而

呈线性形式衰减时(图5-28),其表达式为

$$F_i = f_0(1 - r_i) \qquad (5-30)$$

$$r_i = d_i / d_0 \qquad (5-31)$$

$$0 \leqslant r_i \leqslant 1$$

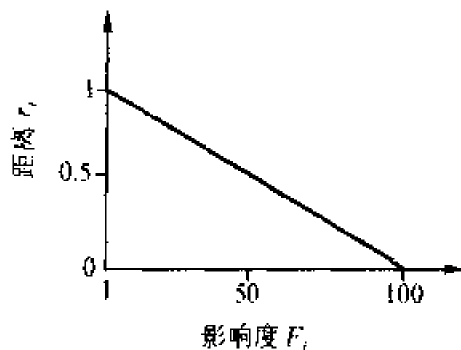


图5-28　线性模型

② 二次模型。用于当主体对邻近对象的影响度(F_i)随距离(r_i)的增大而呈二次形式衰减时(图5-29),其表达式为

$$F_i = f_0(1 - r_i)^2 \qquad (5-32)$$

$$r_i = d_i / d_0 \qquad (5-33)$$

$$0 \leqslant r_i \leqslant 1$$

③ 指数模型。用于当主体对邻近对象的影响度(F_i)随距离(r_i)的增大而呈指数形式衰减时(图5-30),其表达式为

$$F_i = f_0^{(1-r_i)} \qquad (5-34)$$

$$r_i = d_i / d_0$$

$$0 \leqslant r_i \leqslant 1$$

式中:F_i 为主体对邻近对象的实际影响度;

f_0 为主体自身的综合规模指数;

d_i 为邻近对象离开主体的实际距离;

d_0 为主体对邻近对象的最大影响距离。

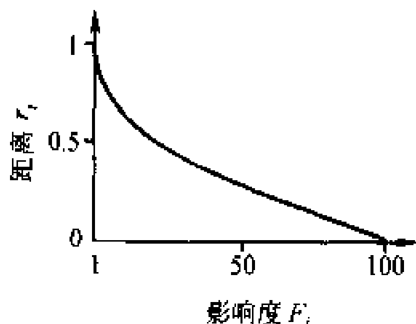


图5-29　二次模型

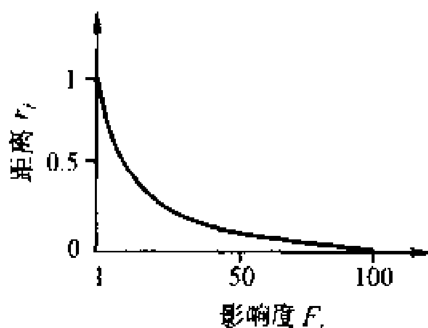


图5-30　指数模型

3.2　空间缓冲区分析的方法

以下,通过实例具体演示空间缓冲区的分析方法和操作过程。

设在某研究区10 km² 区域内有三条道路,它们相关的几何数据和属性数据如表5-10所示。现以这些道路为主体,道路附近的居民出行为邻近对象,试进行这些道路通达度的缓冲区分析。其分析和操作过程如下:

表 5-10 道路数据

名称	坐 标	路宽/m	机动车流量/辆·h^{-1}	非机动车流量/辆·h^{-1}	人流量/人·h^{-1}
A	$x_1y_1,\cdots,x_my_m$	40	182	2 070	2 772
B	$x_1y_1,\cdots,x_ny_n$	22	11	3 991	4 254
C	$x_1y_1,\cdots,x_1y_1$	10	5	725	1 026

1. 计算道路的综合规模标准化指数 f_0

对表 5-10 所列的各项统计数据采用最大值标准化方法，得到表 5-11 的标准化指数 f_0。

表 5-11 处理后的道路数据

名称	路宽	机动车流量	非机动车流量	人流量	综合规模指数	标准化指数 f_0
A	1.00	1.00	0.52	0.65	3.17	100
B	0.55	0.06	1.00	1.00	2.61	82
C	0.25	0.03	0.18	0.24	0.70	22

2. 计算道路的最大影响距离 d_0

道路的最大影响距离 d_0 与该类道路的级别标准和总长度有关，级别标准愈高，则影响距离也愈大，一般按下式推算：

$$d_0 = S/2l \tag{5-35}$$

式中：S 为研究区面积，本例为 10 km^2；

l 为各级道路的长度，分别为 $l_A = 10\,000$ m，$l_B = 4\,286$ m，$l_C = 35\,714$ m，

则道路 A 的 $d_0 = S/2l_A = 500$ m，道路 B 的 $d_0 = S/2(l_A + l_B) = 350$ m，道路 C 的 $d_0 = S/2(l_A + l_B + l_C) = 100$ m。

3. 实施缓冲区操作

道路通达度具有随着离开道路中心线呈迅速递减的特点，因此实施道路通达度的缓冲区操作适宜选择指数形式的分析模型。具体实现的技术途径有以下两种选择：

(1) 由设定 d_i 值→求取 F_i 值→输出缓冲区图形的技术途径。根据该技术途径的操作步骤是，首先，根据应用需求和道路的最大影响距离，分别设定它们的 d_i 值，例如：道路 A 的 d_i 值分别为 100、200、300、400、500(m)；道路 B 的 d_i 值分别为 50、100、150、200、250、300、350(m)；道路 C 的 d_i 值分别为 25、50、75、100(m)。其次，根据式(5-33)和(5-34)分别计算所有道路在不同 d_i 时的 r_i 和 F_i 值。最后，依据 d_i 值在道路的两边绘制平行线，在线的端点处绘制半圆，生成缓

冲区多边形，并在该缓冲区多边形内赋予相应的属性值 F_i，直至输出全部图形及其属性。

(2) 由设定 F_i 值→求取 d_i 值→输出缓冲区图形的技术途径。根据该技术途径的操作步骤是，首先，对式(5－34)作如下变换：

$$d_i = d_0\left(1 - \frac{\ln F_i}{\ln f_0}\right) \tag{5-36}$$

其次，根据应用需求设定 F_i 值，例如：20、40、80、100 等，并利用式(5－36)计算对应的 d_i 值。最后，依据 d_i 值生成道路两边的缓冲区多边形，该缓冲区多边形内部的属性值便与事先设定的需求值相一致，同样直至输出全部图形及其属性。

根据该技术途径输出的点状、线状、面状实体的缓冲区分析图如图 5－31、图 5－32 和图 5－33 所示。

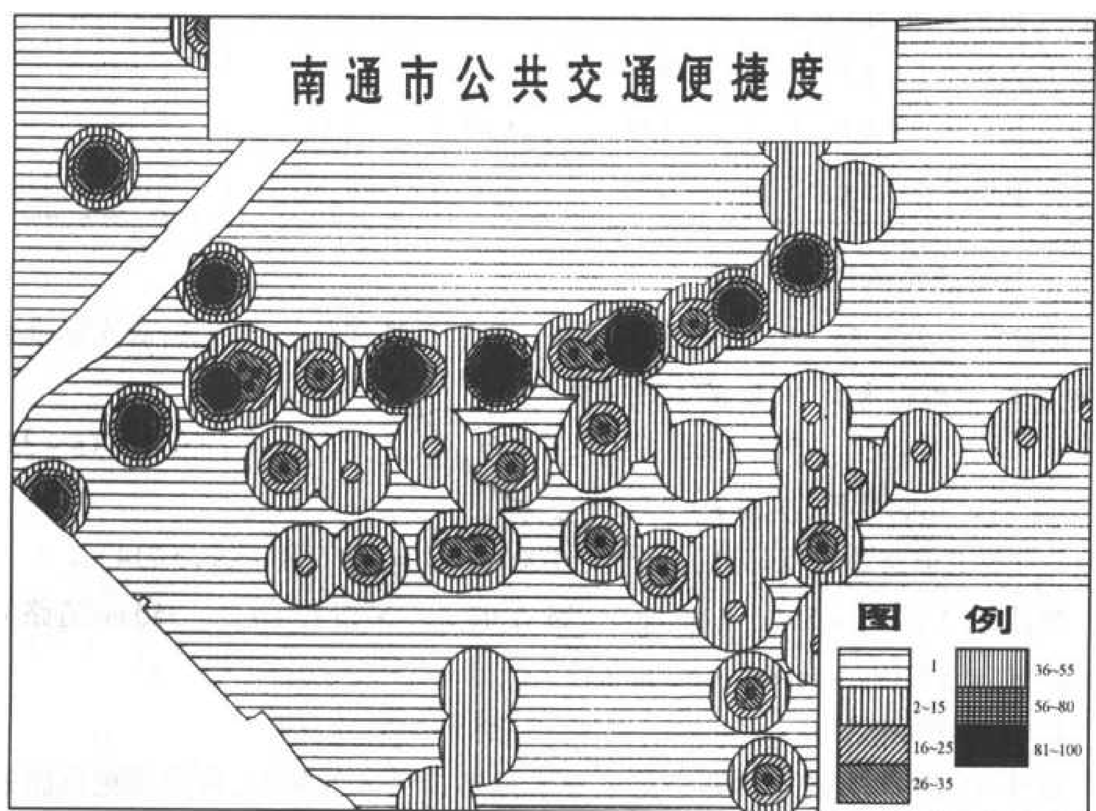


图 5－31　点状实体动态缓冲区分析图

在建立或分析缓冲区时，还应注意到以下几个方面：

① 当缓冲区发生重叠时的处理。缓冲区的重叠包括多个特征缓冲区之间的重叠(图 5－34)和同一特征缓冲区图形的重叠(图 5－35)。对于多个特征缓冲区之间的重叠，首先通过拓扑分析的方法，自动地识别出落在某个缓冲区内部的那些线段或弧段，然后删除这些线段或弧段，得到经处理后的连通缓冲区(图5－34(c))。对于同一特征缓冲区图形的重叠，通过逐条线段求交。如果有交点

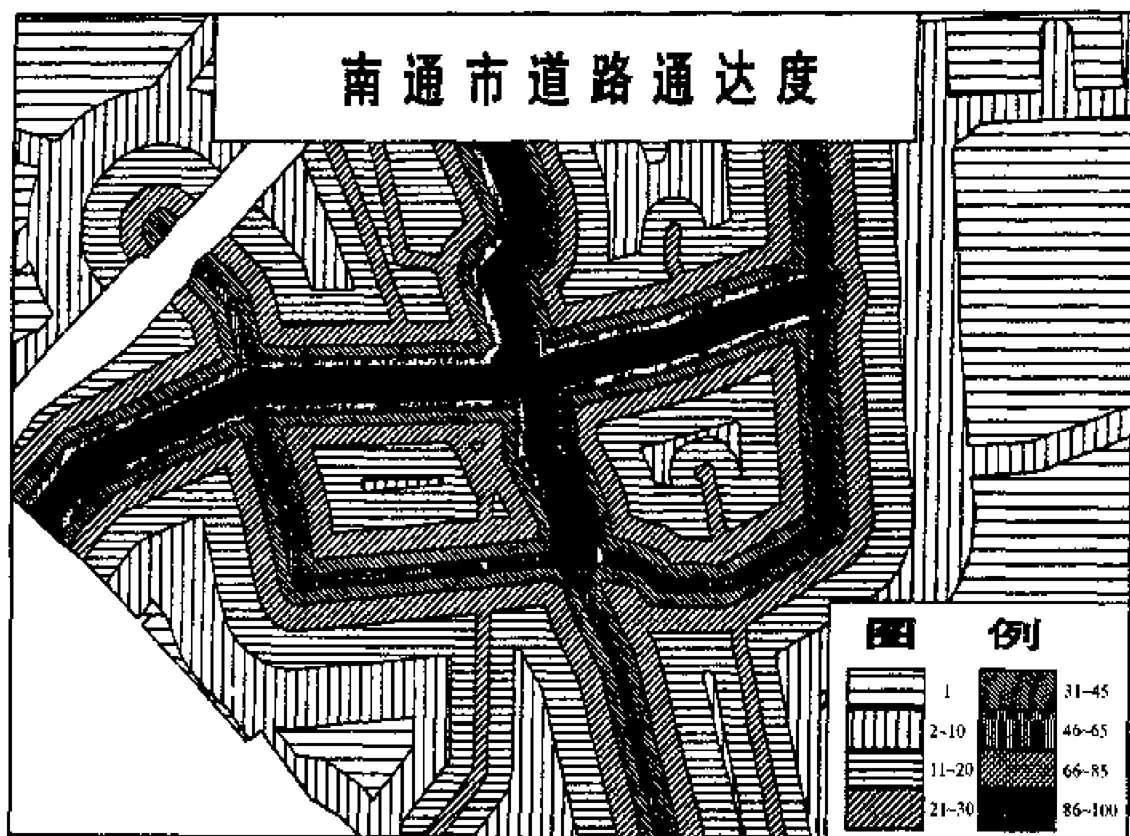


图 5－32 线状实体动态缓冲区分析图

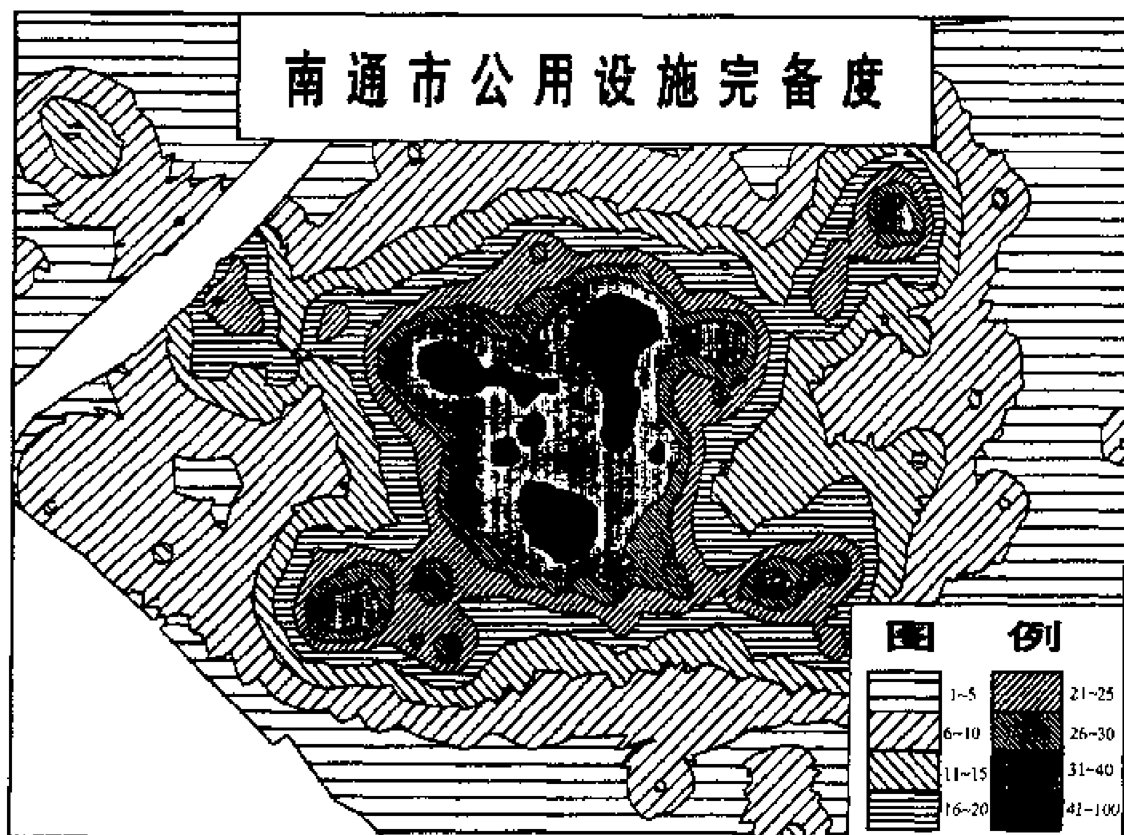
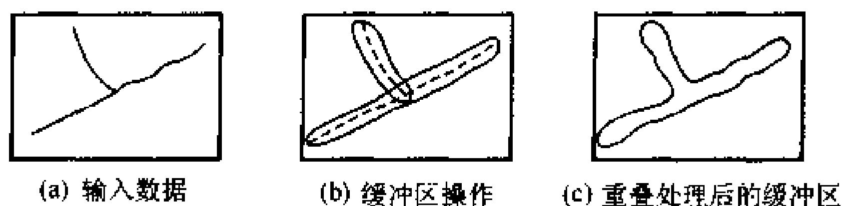


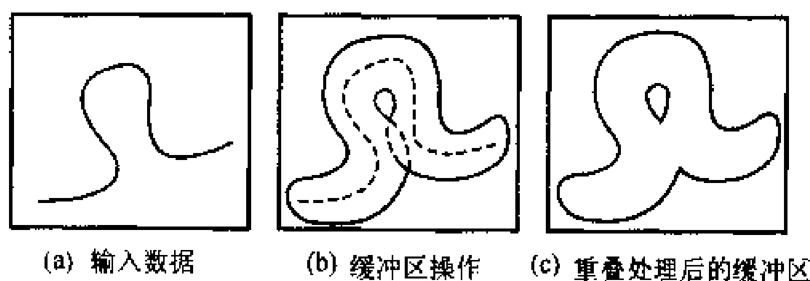
图 5－33 面状实体动态缓冲区分析图



(a) 输入数据　　(b) 缓冲区操作　　(c) 重叠处理后的缓冲区

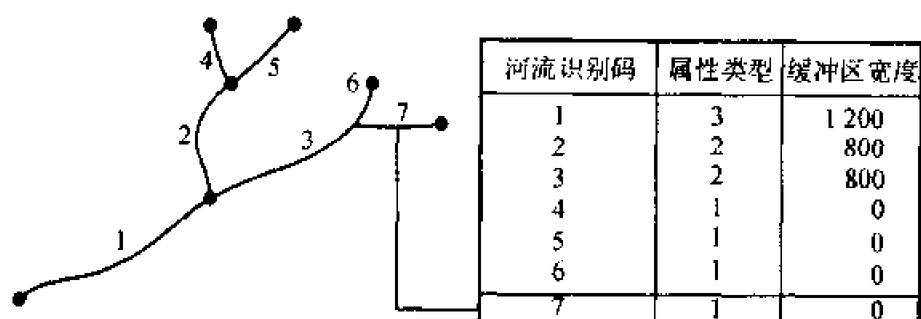
图 5－34　多个特征缓冲区图形的处理

（据 ESRI，1988）



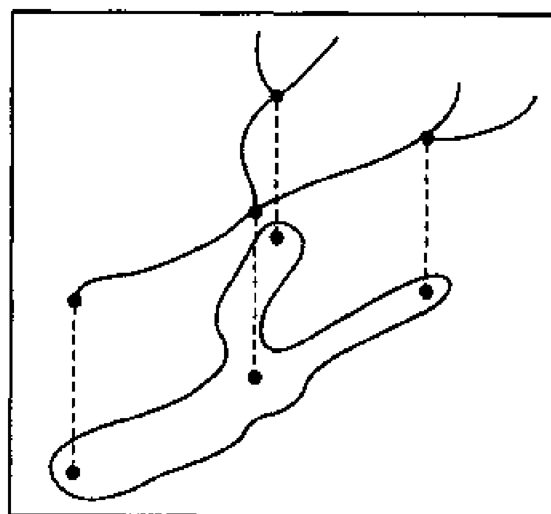
(a) 输入数据　　(b) 缓冲区操作　　(c) 重叠处理后的缓冲区

图 5－35　同一特征缓冲区图形的处理



河流识别码	属性类型	缓冲区宽度
1	3	1 200
2	2	800
3	2	800
4	1	0
5	1	0
6	1	0
7	1	0

(a)



(b)

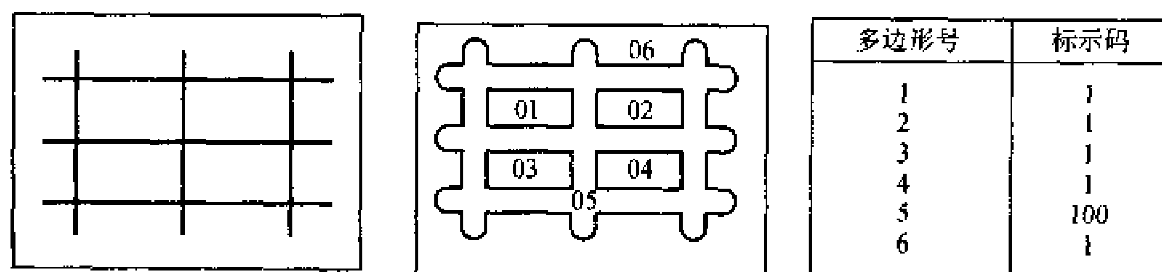
图 5－36　河流类型及其相应宽度缓冲区的建立

（据 ESRI，1988）

且在该两条线段上，则记录该交点。至于该线段的第二个端点是否要保留，则看其是进入重叠区，还是从重叠区出来。对进入重叠区的点予以删除，否则记录之，便得到包括岛状图形的缓冲区(图 5－35(c))。

② 当对特征规定不同缓冲区宽度时的处理。例如，沿河流绘出的环境敏感区的宽度，与各段河流的类型及其特点有关，则通过建立河流属性表，根据不同属性确定其不同的缓冲区宽度(图 5－36(a))，可以产生所需要的缓冲区(图 5－36(b))。

③ 复杂图形情况下的缓冲区与非缓冲区的标示处理。对于原来复杂的要素，经缓冲区分析后，将生成许多的多边形，这些新多边形互相交织在一起，将很难分出哪些是缓冲区，哪些是非缓冲区，这时须对每个多边形加以特征属性的标示。如图 5－37 所示，1 表示属于非缓冲区的多边形，100 表示属于缓冲区的多边形。



多边形号	标示码
1	1
2	1
3	1
4	1
5	100
6	1

图 5－37　缓冲区内外多边形的标示

④ 当需要对缓冲区图形与原输入图形作比较分析时的处理。由于所产生的缓冲区总是一些新的多边形，它们不包括原来输入的点、线、面要素，如果需要显示这些原来的要素，可以利用系统的制图功能(例如 ARC/INFO 系统中的 ARCPLOT)将它们绘出来，以供生成的缓冲区图形与原输入的要素图形比较分析。

§4　空间网络分析

空间网络分析(spatial network analysis)是 GIS 空间分析的重要组成部分。网络是一个由点、线的二元关系构成的系统，通常用来描述某种资源或物质在空间上的运动。城市的道路系统、各类地下管网系统、流域的水网等，都可以用网络来表示，形成各类物质、能量和信息流通的通道。这种面向网络的数据，在 GIS 空间数据中占有较大的比重。

网络分析的用途很广泛，如公共交通运营线路选择和紧急救援行动线路的选择等，与网络最佳路径选择有关；当估计排水系统在暴雨期间是否溢流及河流

是否泛滥时，需要进行网流量分析或负荷估计；城市消防站分布和医疗保健机构的配置等，可以看成是利用网络和相关数据进行资源的分配等等。

面向网络的数据通常用图的形式来模拟，任何一个能用二元关系描述的系统，都可以用图提供数学模型，因此网络图论是空间网络分析的重要理论基础。

4.1 网络图论的基本概念

图论中的“图”并不是通常意义下的几何图形或物体的形状图，而是一个以抽象的形式来表达确定的事物，以及事物之间是否具备某种特定关系的数学系统。

一个图 G 是指由一个非空集合 $V(G)=\{V_i\}$ 和 $V(G)$ 中元素的无序对的一个集合 $E(G)=\{e_k\}$ 所构成的二元组 $(V(G),E(G))$。$V(G)$ 中的元素 V_i 叫做顶点；$E(G)$ 中的元素 e_k 叫做边或弧。

$V=\{V_1,V_2,\cdots,V_n\}$

$E=\{(V_i,V_j)/V_i\in V,V_j\in V\}$

设 $G_d=(V,E)$ 是一个有向图(图5-38)，若 $e_k=V_iV_j\in E$，则称顶点 V_i 和 V_j 是边 e_k 的结点，且称 V_i 是边 e_k 的起始结点，V_j 是边 e_k 的终止结点，并称 V_i 邻接于 V_j 或 V_j 邻接；反之，若移去 G_d 中各边的箭头，则构成无向图。图中两个端点重合的边称为环。两条边以端点相并连接，叫做并行边或重边。若不含并行边和环的图，称为简单图。若 $V(G)$ 和 $E(G)$ 是有限集合，则 G 称为有限图；否则，称为无限图。在无向图中，首尾相接的一串边的集合称为路，如图5-39中的 $\{e_1,e_2,e_3\}$ 构成一条路。能构成自行闭合的路，称为回路。如果任意两个结点之间都存在一条路，称为连通图。若一个连通图中不存在任何回路，则称为树，树中的边，叫做树枝。

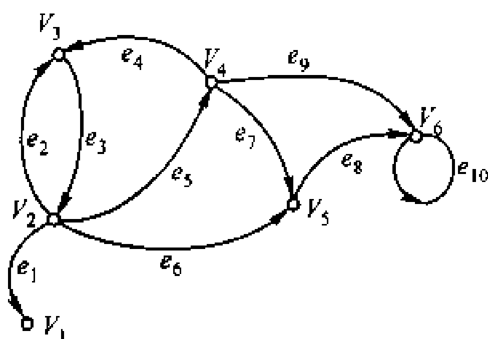


图5-38 有向图 G_d

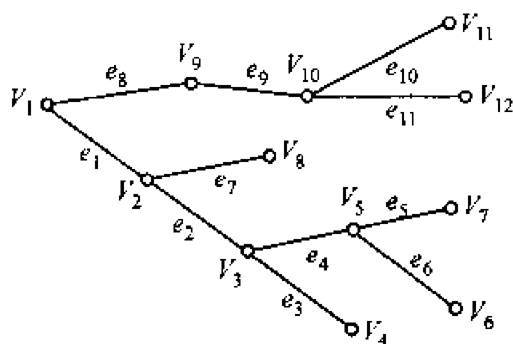


图5-39 树 T

对于给定的有向图 G_d，有时需要对图中任一边或每条弧 e 赋一个实数

$w(e)$，该 $w(e)$ 称为弧 e 的权数。赋权的有向图 G_d，称为赋权有向图，记为 $G_d=(V,E,W)$。

描述一个图的最直观方法是用图形表示，但为了将图输入计算机中，图论中常常使用矩阵来记录图。图的矩阵表示形式很多，其中最基本的矩阵是邻接矩阵 $D(G)$ 和关联矩阵 $A(G)$。

一个具有 V 个顶点、e 条边的无向图 G，可由图 G 的顶点集 V 中每两点间邻接关系惟一决定，其对应的矩阵 $D(G)=[d_{ij}]$ 是一个 $V\times V$ 阶方阵，叫做邻接矩阵。其中

$d_{ij}=1$　　V_i 和 V_j 邻接

$d_{ij}=0$　　V_i 和 V_j 不邻接

（或 $i=j$，且无环）

以图 5-40 为例，其邻接矩阵 $D(G)$ 为

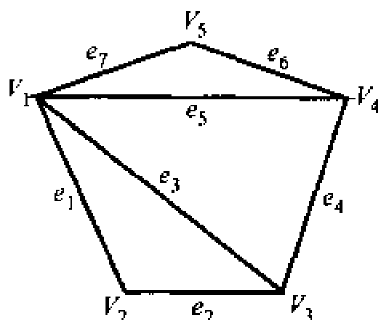


图 5-40　无向图 G 及其邻接矩阵

$$D(G)=\begin{array}{c} \\ v_1\\ v_2\\ v_3\\ v_4\\ v_5\end{array}\begin{array}{c}\begin{array}{ccccc} v_1 & v_2 & v_3 & v_4 & v_5\end{array}\\ \begin{bmatrix}0&1&1&1&1\\1&0&1&0&0\\1&1&0&1&0\\1&0&1&0&1\\1&0&0&1&0\end{bmatrix}\end{array}$$

关联矩阵是图的一个重要矩阵形式，一个具有 v 个顶点和 e 条边的无向图 G 的关联矩阵 $A(G)$，是 $v\times e$ 阶矩阵，每个顶点对应矩阵的一行，每条边对应矩阵的一列，即

$$A(G)=[a_{ij}]_{v\times e}$$

式中：$a_{ij}=1$，当 e_j 和 v_i 关联；

$a_{ij}=0$，当 e_i 和 v_i 不关联。

同样，以图 5-40 为例，其关联矩阵 $A(G)$ 为

$$
A(G)=[a_{ij}]_{5\times 7}=
\begin{array}{c}
 \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5
\end{array}
\begin{array}{c}
\begin{array}{ccccccc} e_1 & e_2 & e_3 & e_4 & e_5 & e_6 & e_7 \end{array} \\
\begin{bmatrix}
1 & 0 & 1 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix}
\end{array}
$$

邻接矩阵和关联矩阵作为图的数字表示形式，在许多使用图的信息系统中发挥了重要作用，但是要使它们能真正适用于 GIS 空间网络的分析，还需要作进一步的研究，有关的具体算法将在空间网络分析方法中介绍。

4.2 空间网络的类型和构成

在地理空间中，由于面向网络的地理目标具有不同的形态，因此构成的空间网络也有着不同的类型。根据空间网络的拓扑学分类，一般可分为平面网络和非平面网络两大类，其基本分类系统如图 5-41 所示。

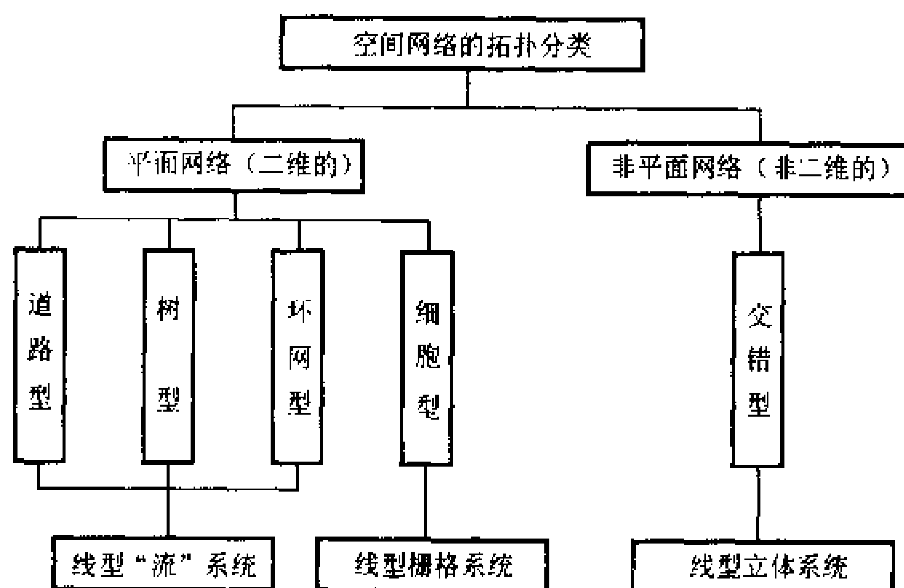


图 5-41 空间网络的拓扑分类
（据牛文元修改）

平面网络中的道路型和树型网络是空间网络中最主要的类型，如交通网抽象后的拓扑结构图就是道路型网络。树型网络最典型的例子是河道的组成，从河源、支流、干流直到河口，构成典型的树型网络。在树型网络中，点和线的功能具有明显的等级特性，这种等级既不容颠倒，也不容穿插。环网型网络是指网络中具有封闭的环形结构，它和树型网络一样，具有引导“流”纳入通道的功能。细胞型网络也称为栅栏状网络，它具有阻断“流”的障碍存在，因此地理学家常利用

这种网络的拓扑分析方法去研究行政区划系统和土地利用系统。交错型网络的典型例子是城市地下管线网,包括给水、排水、电力、电信、煤气、热力、工业等多类别及多权属和布局复杂的管线网,这类网络的主要特征是具有复杂的横断面和纵剖面结构。

空间网络除具有一般网络的边、结点间抽象的拓扑特征之外,还具有 GIS 空间数据的几何定位特征和地理属性特征。各类空间网络虽然形态各异,但是构成网络的基本元素(图 5-42)主要包括:

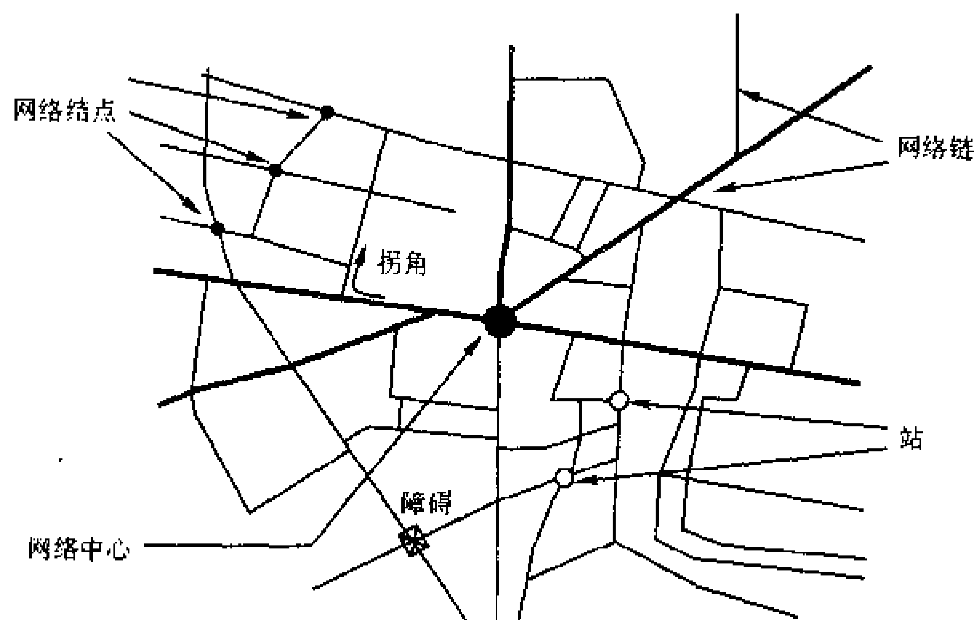


图 5-42 空间网络的构成元素

(据白玲)

(1) 结点。网络中任意两条线段或路径的交点,其属性如方向数、资源数量等。

(2) 链或弧段。连接两个结点的弧段或路径,是网络中资源运移的通道。其属性如资源流动的时间、速度、资源种类和数量、弧段长度等。

(3) 障碍。指资源不能通过的结点,如被破坏的桥梁和禁止通行的关口等。它是惟一不表示任何属性的元素。

(4) 拐角。在网络的结点处,资源运移方向可能转变,从一个链经结点转向另一个链,例如在十字路口禁止车辆左拐,便构成拐角。其拐角的类型及其属性描述见表 5-12。

(5) 中心。指网络中具有从链上接受或发送资源能力的结点所在地,如水库属于河网的中心,学校属于路网的中心等等。其属性如资源最大容量、最大服务半径等。

(6) 站点。是网络中装卸资源的结点所在地,例如车站、码头等。其属性如

资源需求量,正值表示装载量,负值表示下卸量。

表5-12 拐角的类型及其属性描述

转弯类型	描述	属性表 0=无阻强 -1=不允许拐弯				
		结点号	从弧段	至弧段	角度	时间阻强/s
U型拐弯	U型拐弯指从6号弧至20号结点并从20号结点转回6号弧,这是一个180度转弯,花费20秒时间	20	6	6	180	20
停靠点	停靠点使得从6号弧至其他弧段——直通7号弧,向左转至8号弧,向右转至9号弧的运移减慢	20	6	7	0	15
		20	6	8	90	20
		20	6	9	-90	10
高架道或地道	高架道或地道允许直通而无延迟,如从6号弧至7号弧;但不允许转弯,此时以负的阻强表示,如从6号弧至8、9号弧	20	6	7	0	0
		20	6	8	90	-1
		20	6	9	-90	-1
		20	9	8	0	0
		20	9	7	-90	-1
		20	9	6	90	-1
不准右转弯	不允许从6号弧转向9号弧,并赋予负值阻强;允许其他方向的转变,其阻强为正	20	6	9	-90	-1
		20	6	7	0	5
		20	6	8	90	10

注:据郭达志等,1997。

4.3 空间网络分析方法

1. 路径分析(path analysis)

在空间网络分析中,路径问题占有重要位置。人们常想知道在地理空间网络的两指定结点间是否存在路径,如果有则特别希望找出其中的最短路径。这种路径问题对于交通、消防、信息传输、救灾、抢险等有着重要的意义。在运输网络中,有时要找出运输费用最小的路径;在通讯网络中,要找出两点间进行信息传递具有最大可靠性的路径等等。由于大量的最优化问题等价于找一个网络图的最短路径的问题,因而引起了人们对于最短路径分析的极大兴趣。下面介绍

的最短路径搜索算法是 Dijkstra 在 1959 年提出的，被公认为是最好的算法之一。

为了求出最短路径，需先计算网络任意两点间的距离，并形成 $n \times n$ 阶距离矩阵或权矩阵。

$$W = [W_{ij}]$$

式中：w_{ij} 为网络中的边 e_{ij} 的距离。

在矩阵 W 中，

$w_{ij} > 0$，当 i, j 间有边相连接时，对于无向图，$w_{ij} = w_{ji} (i \neq j)$；

$w_{ij} = \infty$，当 i, j 间无边相连接时；

$w_{ij} = 0$，当 $i = j$ 时。

图 5－43 给出一无向图 G，它的距离矩阵 W 如式(5－37)所示。

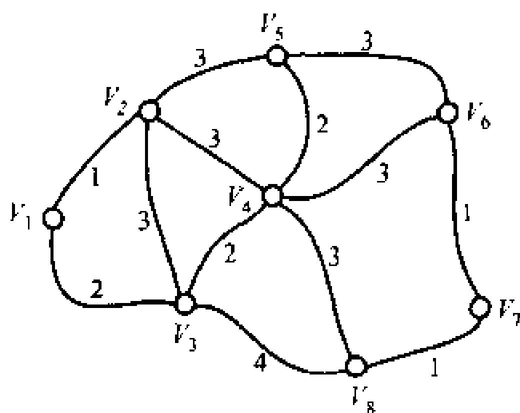


图 5－43　加权无向图 G(据陈树柏等)

$$W = \begin{array}{c} \\ v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \\ v_8 \end{array} \begin{array}{c} \begin{array}{cccccccc} v_1 & v_2 & v_3 & v_4 & v_5 & v_6 & v_7 & v_8 \end{array} \\ \begin{bmatrix} 0 & 1 & 2 & \infty & \infty & \infty & \infty & \infty \\ 1 & 0 & 3 & 3 & 3 & \infty & \infty & \infty \\ 2 & 3 & 0 & 2 & \infty & \infty & \infty & 4 \\ \infty & 3 & 2 & 0 & 2 & 3 & \infty & 3 \\ \infty & 3 & \infty & 2 & 0 & 3 & \infty & \infty \\ \infty & \infty & \infty & 3 & 3 & 0 & 1 & \infty \\ \infty & \infty & \infty & \infty & \infty & 1 & 0 & 1 \\ \infty & \infty & 4 & 3 & \infty & \infty & 1 & 0 \end{bmatrix} \end{array} \tag{5-37}$$

DijKstra 算法是一种对结点不断进行标号的算法。每次标号一个结点，标号的值即为从给定起点到该点的最短路径长度。在标定一个结点的同时，还对所有未标号结点给出了“暂时标号”即当时能够确定的相对最小值。设定 K 表示待确定最短路径的起点，L 表示终点，则最短路径搜索的步骤如下：

(1) 令起点 K 标号为零，其他结点标号为 ∞。

(2) 对未被定标的结点全部给出暂时标号,其值为 min[j 的旧标号,(i 的旧标号 + w_{ij})],这里 i 是前一步刚被标定的结点,w_{ij} 是边 e_{ij} 的权,如果结点 i 和 j 不相邻接,$w_{ij} = \infty$。

(3) 找出所有暂时标号的最小值,用它作为相应结点的固定标号。如果存在几个有同一最小标号值的结点,则可任取一个加以定标。

(4) 重复进行(2)与(3),直至指定的终点 L 被定标时为止。用此法可直接得到由起点 K 到其他结点的最短路径的长度,那就是该结点的定标数值。

对于图 5-43 中的加权无向图 G,从结点 v_1 到 v_7 的最短路径的标号过程如下:

	v_1	v_2	v_3	v_4	v_5	v_6	v_7	v_8
确定起点	0	(∞)	(∞)	(∞)	(∞)	(∞)	(∞)	(∞)
标定 v_1	0	(1)	(2)	(∞)	(∞)	(∞)	(∞)	(∞)
标定 v_2	0	1	(2)	(4)	(4)	(∞)	(∞)	(∞)
标定 v_3	0	1	2	(4)	(4)	(∞)	(∞)	(6)
标定 v_4	0	1	2	4	(4)	(7)	(∞)	(6)
标定 v_5	0	1	2	4	4	(7)	(∞)	(6)
标定 v_8	0	1	2	4	4	(7)	(7)	6
标定 v_7	0	1	2	4	4	(7)	7	6

其中括号内的是暂时标号,没有括号的为定标。距离起点越近的顶点,越早得到固定标号。采用回溯的方法可以得到从起点 K 到其他结点的最短路径经由的结点。因此,从结点 v_1 到 v_7 的最短路径长度为 7,经由的路径为 $v_1 \to v_3 \to v_8 \to v_7$。

2. 定位-配置分析(location-allocation analysis)

定位-配置分析是指根据中心地理论框架,通过对供给系统和需求系统两者空间行为相互作用的分析,来实现网络设施布局的最优化。其中,若已设定需求点,求供给点,则涉及定位问题(location);若已设定供给点,求需求分配点,则涉及配置问题(allocation);若同时求供给点和需求分配点,则涉及定位-配置问题(location-allocation)。

这类问题在城市与区域规划中应用非常广泛,例如选择最佳布局中心,或者从一批候选位置中选定若干地点来建设公共设施,为区域的需求点提供服务。这些服务设施诸如医疗保健、邮电通讯、交通站点、行政中心等,它们是城市或区域规划的重要内容。定位-配置分析也是 GIS 空间分析研究的热点之一。

定位－配置分析涉及的因素比较多，除了网络的元素数据，还包括规划时间范围、问题空间类型、公共设施服务方式、费用对象、设施使用类型、需求点分配类型、附属区域等。因此，当进行定位－配置分析时，首先要建立一系列边界条件和确定若干目标函数。边界条件代表了规划目标所必须满足的规划条件，例如，要求所有需求点都有相应的供给点等，作为问题解决的约束条件；目标函数是给出极大值或极小值，例如规定所有设施和需求点之间的总加权距离为最小等，其意义是在于能获得一个明确的分析结果。

定位－配置分析的算法包括 P—中心问题、中心服务范围的确定和中心资源的分配范围等，它们是定位－配置分析或规划的重要工具，以下分别予以介绍。①

(1) P—中心问题。这一问题是要在 m 个候选点中，选择 P 个供应点，为 n 个需求点服务，并使得从服务中心到需求点之间的总距离(或时间、费用)为最小。P—中心问题可以描述为：

$$\min\left(\sum_{i=1}^{n}\sum_{j=1}^{m} w_i \cdot d_{ij} \cdot x_{ij}\right)$$

并满足 $\sum_{j=1}^{m} x_{ij} = 1$ 保证每个需求点 $i(i = 1,2,\cdots,n)$ 都被服务

和 $\sum_{j=1}^{m} y = P$ 将服务中心的数量限定为 P 个。

上述两个约束条件是为了保证每个需求点仅受一个供应点服务，并且只有 P 个供应点。

式中：i、n 分别为需求点的位置和数量；

j、m 为候选点的位置和数量；

P 为要确定的服务中心的数量；

w_i 为需求点 i 的需求量；

d_{ij}为需求点 i 到服务点 j 的距离。

并且 $y_i \geqslant x_{ij}, \forall_i, \forall_j$ 只有第 j 个中心被选中时，需求点 i 才能分配到该中心；

$y_j = 0$ 或 $1, \forall_j$ 任何一个候选点被选中时为 1；否则为 0；

$x_{ij} = 0, \forall_i, \forall_j$ 需求点有中心 j 服务时为 1；否则为 0。

P—中心问题可以用线性规划方法求得全局性的最佳结果，但由于计算量及内存需求量巨大，在实际应用中常用一些启发式算法来逼近或求得最佳结果，其中著名的有 Teitz－Bart 算法。该算法的主要步骤如下：

① 以下算法主要参照学位论文《GIS 中地理网络分析的研究与实践》(白玲)写成。

① 先选 P 个候选点作为起始供应点集 $P_t: C_1, C_2, \cdots, C_P$。

② 将所有的需求点分配给它们最邻近的供应点，使其距离为最短。计算总的加权距离为 B_t。

③ 从未被选取的候选点集中选一候选点 C_b。

④ 对 P_t 中的每个供应点 C_j 用 C_b 替换之，并计算其总加权距离的变化 Δb_j。

⑤ 如果用 C_b 替换某个 C_j 后，可以使总加权距离减少，那么就替换总加权距离减少最多的供应点 C_k，令 $B_t = B_t - \Delta b_k$，并将 P_t 修改为 B_t 所在的供应点。

⑥ 重复③～⑤步，直至未被选取的候选点集为空。

⑦ 当所有不在 P_t 中的候选点都试过后，其结果记为 P'_t，并取代 P_t。继续重复②～⑦步，如果没有任何取代能减少总的加权距离，则停止。其最后的结果 P'_t 即为所求的 P 个中心的供应点。

(2) 中心服务范围的确定。中心服务范围是指一个服务设施在给定的时间或距离内，能够到达的区域。

设一个带中心的空间网络 $G=(V,E,C)$，其中 V 表示空间网络结点的集合，E 表示边的集合，C 为该网络的一个中心。若已知该中心的阻值为 cw，网络边 e_{ij} 的费用为 w_{ij}，r 表示空间网络上任何结点到中心的 (v_i, v_c) 间的一条路径，r_{ic} 是该路径的费用，那么在不考虑货源量和需求量的情况下，中心的服务范围应为满足下列条件的网络边和结点的集合 F：

$$F = \{v_i \mid r_{ic} \leqslant cw, v_i \in r\} \cup \{e_{ij} \mid r_{ic} + w_{ij} \leqslant cw, v_i \in r\}$$

为确定该中心的服务范围，须依次求出到服务中心费用不超过中心最大阻值的路径，于是组成这些路径的网络结点和边的集合，就构成该中心的服务范围。具体处理时，将空间网络看作是以该中心为根的一棵树，用宽度优先法搜索，搜索时需考虑到网络边和结点的属性。其具体的算法步骤如下：

① 将中心结点作为当前结点，并存入到达结点表中，同时扫描它的关联结点，检查这些结点是否在服务范围内，将其中满足条件的关联结点存入搜索结点表中。

② 如果已搜索结点表为空，则结束；否则，继续第③步。

③ 在已搜索结点表中，按表中记录的顺序取出关联结点，作为当前已到达结点，存入已检查到达结点表中，同时记录该结点所在的边。

④ 扫描当前结点的关联结点，检查这些结点是否在服务范围内，将满足条件的关联结点存入已搜索结点表中。

⑤ 重复执行第②～④步。

求解中心服务范围的程序框图如图 5-44 所示。

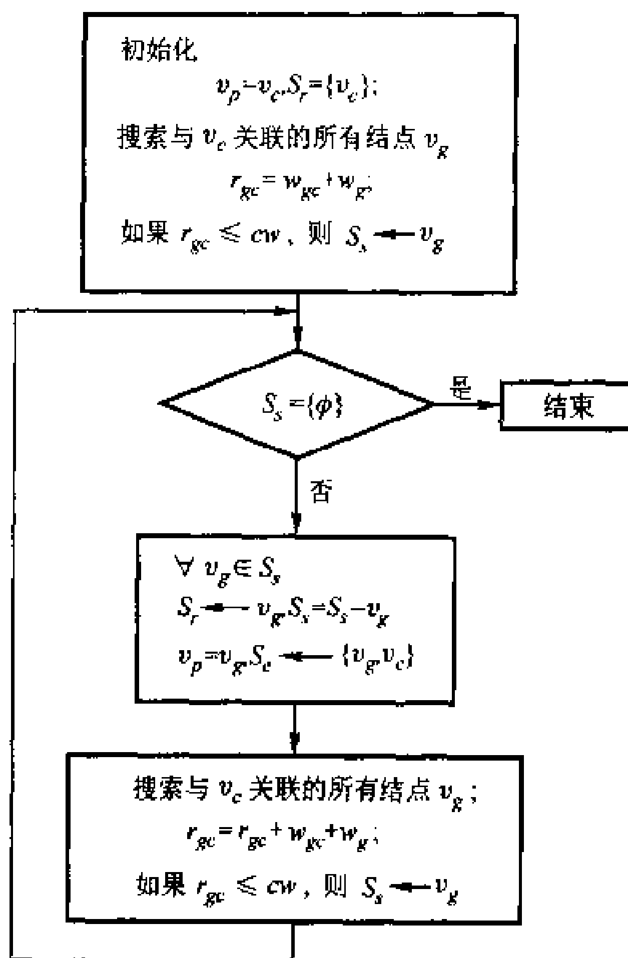


图 5-44 求解中心服务范围的程序框图

(3) 中心资源的分配范围。资源分配就是将空间网络的边或结点，按照中心的供应量及网络边和结点的需求量，分配给一个中心的过程，它用来模拟空间网络上资源的供需关系。

设一个带中心的空间网络 $G=(V,E,C)$，其中 V 表示空间网络结点的集合，E 表示网络边的集合，C 为网络的一个中心。已知中心的货源量 CS，中心的阻值 cw，d_{ij} 和 w_{ij} 分别表示网络边 e_{ij} 的需求量和费用，r 表示网络上任何结点到中心的 (v_i,v_c) 间的最短路径，r_{ic} 是该路径的费用，m 是网络当前的总需求量。则资源的分配范围定义为满足下列条件的网络边和结点的集合 P：

$$P=\{v_i \mid r_{ic}\leqslant cw, m\leqslant cs, v_i\in r\}$$
$$\cup\{e_{ij}\mid r_{ic}+w_{ij}\leqslant cw, m+w_{ij}\leqslant cs, v_i\in r\}$$

资源分配范围的求解思想是，依次求出到服务中心费用不超过中心最大阻值，同时网络的总需求量不超过中心的货源量的路径，组成这些路径的网络结点和边的集合就构成了该中心资源分配的范围。处理时同时考虑到网络和网络结点的

需求量，具体处理时与服务范围的搜索方式类似。

算法的主要步骤如下：

① (初始化)将中心所在的结点作为 v_0，存入已标记结点集 S 中，并初始化有关变量。

② 如果整个网络都已被分配，则停止；否则，执行第①步。

③ 如果总货源量都已被分配，则停止；否则，执行第(4)步。

④ 在尚未分配的结点集 $\bar{S}$ 中，寻找距离中心 v_0 路径最短的结点 v_n，假设 v_n 的前一点是 v_m，将 (v_m, v_n) 作为当前处理的边。

⑤ 判断网络流在边 (v_m, v_n) 上的运行情况：

边接受到来自 v_m 点的流量：$LR_{mn} = \min\{LL_{mn}, PO_m\}$

边消耗掉的流量：$LF_{mn} = \min\{LD_{mn}, LR_{mn}\}$

由该边流向 v_n 的流量：$LO_{mn} = LR_{mn} - LF_{mn}$

其中，LD_{mn}，LL_{mn} 分别为 (v_m, v_n) 边的需求量和通行能力；PO_m 为 v_m 点发出的流量。

⑥ 如果 (v_m, v_n) 边流向 v_n 点的流量为0，则该边停止运输。

如果 (v_m, v_n) 边流向 v_n 点的流量小于该边的需求量，则将该边的一部分分配给中心后，停止运输。

如果 (v_m, v_n) 边流向 v_n 点的流量大于该边的需求量，则考察网络流在 v_n 点上的接受量 PR_n、消耗量 PF_n 和发出量 PO_n。

⑦ 判断网络流在结点 v_n 上的运行情况，与网络流在边上的运行情况类似：

结点 v_n 接受到来自 (v_m, v_n) 的流量：$PR_n = \min\{PL_n, LO_{mn}\}$

结点 v_n 消耗掉的流量：$PF_n = \min\{PD_n, PR_n\}$

由结点 v_n 流向相邻边的流量：$PO_n = PR_n - PF_n$

其中，PD_n，PL_n 分别为结点 v_n 的需求量和通行能力。

⑧ 如果 $PO_n \leq 0$，则该点停止运输。

如果 $PO_n > 0$，则考察与结点 v_n 相邻的边。

⑨ 如果存在与 v_n 点相邻的边 (v_n, v_r)，该边尚未分配而该边的 v_r 点已经分配，则给该边分配它所需要的量 LD_{nr}，此时，从 v_n 点流向其它相邻的，且另一端点尚未分配的边的流量为 $PO_n = PO_n - LD_{nr}$。

⑩ 记录已分配的结点 v_n、边 (v_m, v_n) 或边 (v_n, v_r)，并从未分配的点、边集合 $\bar{S}$ 和 $\bar{Q}$ 中减去这些元素，将点 v_n 作为当前结点。

⑪ 计算全网络点或边的总消耗量 LFT、PFT，点或边的分配数 LNT、PNT 以及总的消耗量 TF。

§5 空间统计分析

空间统计分析主要用于数据分类,而数据分类方法是任何地理信息系统重要的组成部分。一般地,地图上载负的数据是经过专门分类和处理的,而地理信息系统存储的数据具有原始的性质,以使用户可以根据不同的使用目的,进行任意提取和分析,特别是对于观测和取样数据,随着采用的分类和内插方法的不同,得到的结果有很大的差异。因此,在大多数情况下,首先是将大量未经分类的数据输入信息系统的数据库,然后要求用户建立具体的分类算法,以获得所需的信息。

5.1 变量筛选分析

随着现代数据收集系统的不断改进,在一个取样点上常可以收集到几十种原始变量。例如,我国 1982 年的人口普查,每个县都包括有 40 多种的人口统计变量。根据统计学原理,这些复杂的多变量之间有许多是相互关联的,可以通过寻找一组相互独立的变量,使多变量数据得到简化,这就叫变量筛选分析。常用的变量筛选分析方法有主成分分析法、主因子分析法、关键变量分析法等。

主成分分析是以取样点作为坐标轴,以变量作为矢量,通过以相似系数建立相关矩阵,研究变量之间的亲疏关系。主因子分析法是以变量作为坐标轴,以取样点作为矢量,通过以相关系数建立相关矩阵,研究取样点之间的亲疏关系。关键变量分析则是利用变量之间的相关矩阵,通过由用户确定的阈值,从数据库变量全集合中选择一定数量的关键独立变量,以消除其他冗余的变量。其计算步骤如下:

设有 n 个数据点,每点有 m 个变量,则多变量数据可以表示为

$$x = \begin{pmatrix} x_{11} & x_{22} \cdots x_{1m} \\ x_{21} & x_{22} \cdots x_{2m} \\ \vdots & \vdots \quad \vdots \\ x_{n1} & x_{n2} \cdots x_{nm} \end{pmatrix}$$

这些变量之间的关系可用相关系数表示,它是将数据标准化(标准差标准化)后的夹角余弦。其相关系数用 r_{ij} 表示。

$$r_{ij} = \frac{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)(x_{ki} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki} - \bar{x}_i)^2 \sum_{k=1}^{n}(x_{ki} - \bar{x})^2}}$$

式中:i、$j = 1,2,\cdots,m$;

$k = 1,2,\cdots,n$。

r_{ij}的取值范围是$0 \leqslant r_{ij} \leqslant 1$。显然，$|r_{ij}|$越近于1，变量$i$与变量$j$的关系越密切，$r_{ij}$越近于零，两者的关系越疏远。根据$r_{ij}$建立的相关矩阵如表5-13所示，相关阈的值为$t$(例如可定0.81、0.64、0.49、0.36、0.35…)。则关键变量的分析过程为：

(1) 将相关矩阵上角线($j > i$)的所有元素r_{ij}的值取平方$r_{ij} = r_{ij}^2$；

(2) 从相关矩阵中，选取r_{ij}值为最小的两个变量i和j作为两个独立的关键变量；

(3) 其他与i、j之间有联系，且相关系数的平方值大于t的所有变量均从变量表中排除出去；

(4) 其他与i、j之间有联系，且相关系数的平方值为最小的变量，选作独立的关键变量。同时将与该变量有联系，且相关系数的平方值大于t的变量从变量表中加以排除；

(5) 依次继续下去，直到全部变量经过处理，或者关键变量已满足所需要的个数L，则过程结束。例如，根据表5-13建立的相关矩阵，当$t = 0.25$(相当于表5-13的绝对相关系数值0.5)时，得到的关键变量依次为：8、14、7、1、4、5、13、18、10、12和15。其关键变量的具体算法流程图如图5-45所示。

5.2　变量聚类分析

所谓变量聚类分析就是将一组数据点或变量，按照其在性质上亲疏远近的程度进行分类。两个数据点在m维空间的相似性可以用这些点在变量空间的距离(d_{ij})来度量。

$$d_{ij} = \left[\sum_{k=1}^{n}(x_{ik} - x_{jk})^2\right]^{1/2}$$

这里d_{ij}称做欧几里德距离(Euclidean Distance)，它表示数据点i与数据点j之间的距离，x为相应点的变量k数据。距离越小，表明两者的相似性越大。或者采用马氏距离(Mahalonobis's)，数据点i与数据点j的马氏距离(d^2)是

$$d_{ij} = (x_i - x_j)' \sum^{-1}(x_i - x_j)$$

式中：x_i、x_j($i, j = 1, 2, \cdots, n$)为数据点i和j的m个变量所组成的向量；

$\sum^{-1}$表示逆协差阵。

以下，设根据欧几里德距离公式建立距离系数矩阵(表5-14)，利用最短距离法获得聚类图(图5-46)。其算法步骤如下：

(1) 计算数据点之间的距离，形成距离系数矩阵$D(0)$(表5-14)。这时各个数据点自成一类，显然有$D_{pq} = d_{pq}$。

(2)选择$D(0)$中的非对角最小元素，设为D_{pq}，则将G_p和G_q并为一类，记

表 5-13 相关矩阵

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	1																			
2	0.419 7	1																		
3	-0.055 5	0.012 5	1																	
4	0.014 6	0.244 3	0.552 0	1																
5	-0.128 6	-0.125 8	0.652 1	0.001 7	1															
6	0.914 8	0.479 4	-0.001 3	0.089 7	-0.133 6	1														
7	-0.127 2	-0.290 1	0.092 3	-0.162 6	0.091 2	-0.129 5	1													
8	-0.021 6	-0.268 1	-0.452 7	-0.443 8	-0.118 2	-0.113 1	-0.376 3	1												
9	-0.095 5	-0.510 2	-0.084 6	-0.374 1	-0.146 9	-0.194 5	0.749 5	0.053 6	1											
10	-0.198 3	-0.416 3	0.371 2	0.349 9	0.088 6	-0.155 2	0.092 4	-0.334 6	0.172 9	1										
11	-0.141 9	-0.140 1	0.275 8	0.130 3	0.123 5	-0.134 1	0.762 1	-0.602 1	0.394 8	0.274 2	1									
12	-0.050 5	-0.027 8	0.272 1	0.463 0	0.087 2	-0.006 7	-0.404 5	-0.236 1	-0.409 9	0.368 6	0.067 3	1								
13	0.463 2	0.585 4	0.181 3	0.011 9	-0.042 5	0.493 2	-0.329 5	0.144 3	-0.366 3	-0.258 4	-0.197 3	-0.166 1	1							
14	-0.198 9	0.028 8	-0.572 8	-0.174 7	-0.386 2	-0.274 4	0.007 5	-0.000 4	-0.036 2	-0.211 7	0.247	-0.172 7	-0.411 0	1						
15	-0.090 0	0.269 9	-0.264 7	0.141 7	0.255 0	-0.011 9	0.061 3	-0.287 3	-0.111 3	0.062 2	0.186 5	0.201 5	0.209 2	-0.298 6	1					
16	-0.157 8	-0.460 1	-0.414 6	-0.162 4	-0.378 1	-0.153 8	0.132 5	0.286 5	0.115 2	0.106 9	0.198 2	0.081 4	-0.106 9	0.108 2	-0.198 4	1				
17	0.161 2	0.400 5	0.385 5	0.210 8	0.374 2	0.148 1	-0.116 2	-0.312 6	-0.026 2	-0.016 4	-0.160 2	0.085 3	-0.030 9	-0.127	0.118 7	-0.935 7	1			
18	-0.020 1	-0.338 8	-0.479 2	-0.367 9	-0.235 2	-0.090 2	0.198 7	0.214 5	0.324 5	-0.049 2	0.028 0	0.298 7	-0.121 9	0.182 3	-0.268 6	0.561 4	-0.539 6	1		
19	0.229 1	0.640 0	0.593 5	0.410 0	0.201 9	0.334 5	0.070 2	-0.527 3	-0.325 4	-0.069 2	0.218 9	0.043 7	0.432 9	-0.339 3	0.444 0	-0.560 3	0.570 3	-0.574 9	1	
20	0.272 4	0.573 9	0.630 6	0.454 3	0.186 0	0.382 7	0.043 5	-0.488 4	-0.326 5	0.009 9	0.243 6	0.103 1	0.459 9	-0.429 6	0.415 1	-0.474 4	0.392 6	-0.528 6	0.953 5	1

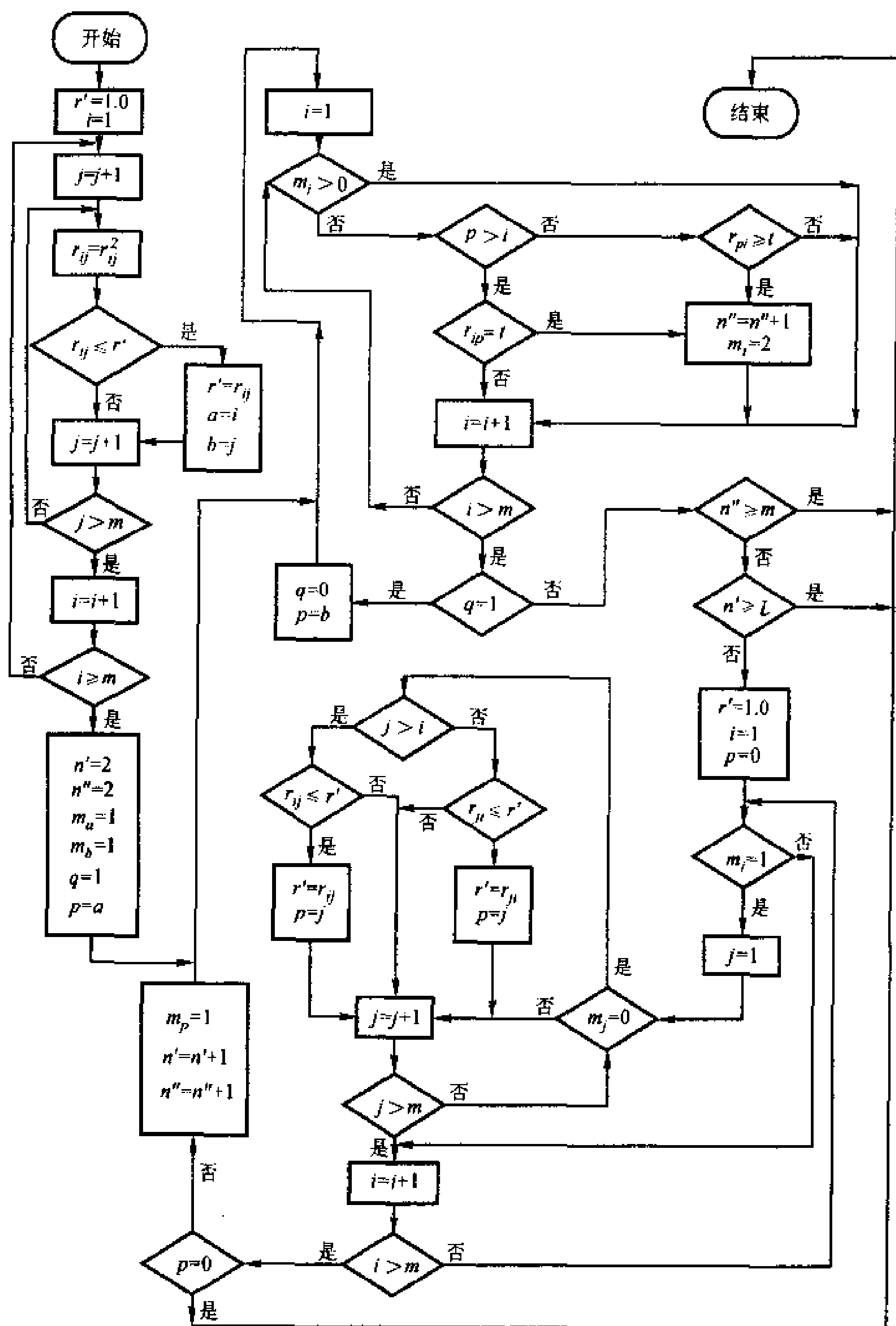


图5-45 关键变量算法流程图

(据 M.S.Monmonier,1978)

表 5-14 距离系数矩阵

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	0																				
2	1.82																				
3	3.22	1.35	0																		
4	2.37	0.51	0.91	0																	
5	3.00	1.18	0.41	0.73	0																
6	3.58	1.66	0.33	1.23	0.72	0															
7	4.23	2.49	1.18	1.98	1.31	0.89	0														
8	3.56	1.82	0.51	1.31	0.74	0.32	0.66	0													
9	1.64	0.52	1.95	1.05	1.76	2.28	2.65	2.06	0												
10	4.04	2.16	0.83	1.69	1.22	0.54	0.35	0.78	2.74	0											
11	3.11	1.27	0.20	0.78	0.33	0.51	1.32	0.65	1.8	1.01	0										
12	3.56	1.64	0.31	1.21	0.75	0.08	0.97	0.30	2.26	0.60	0.45	0									
13	4.18	2.34	1.03	1.85	1.40	0.74	0.33	0.76	2.72	0.20	1.17	0.82	0								
14	2.06	1.10	2.43	1.53	2.08	2.76	2.93	2.24	0.48	2.82	2.29	2.74	3.20	0							
15	1.79	0.27	1.60	0.70	1.29	1.93	2.44	1.79	0.47	2.39	1.46	1.91	2.45	0.83	0						
16	3.95	2.17	0.86	1.66	1.17	0.57	0.52	0.43	2.33	0.67	1.00	1.70	0.65	2.55	2.22	0					
17	2.89	1.15	0.78	0.64	0.39	1.11	1.60	1.05	1.65	1.57	0.64	1.09	1.21	1.97	1.18	1.48	0				
18	2.94	1.20	0.91	0.69	0.66	1.24	1.51	0.95	1.36	1.70	0.71	1.22	1.68	1.52	1.71	1.53	0.44	0			
19	2.84	1.10	0.83	0.69	0.54	1.16	1.39	0.84	1.26	1.62	0.68	1.14	1.60	1.66	1.07	1.15	0.43	0.14	0		
20	3.45	1.67	0.38	1.16	0.67	0.72	1.02	0.47	1.83	1.17	0.49	0.69	1.15	2.05	1.72	0.50	0.93	0.53	0.65	0	
21	3.22	1.38	0.59	0.89	0.46	0.92	1.31	0.76	1.58	1.38	0.45	0.90	1.38	1.83	1.49	0.81	0.75	0.32	0.42	0.31	0

为 G_r，$G_r=\{G_p,G_q\}$。

根据表5-14，$G_r=\{G_6,G_{12}\}$。

(3) 计算新类与其他类的距离

$$
\begin{aligned}
D_{rk} &= \min_{i \ni G_r,\, j \ni G_k} d_{ij} \\
&= \min\{\min_{i \ni G_p,\, j \ni G_k} d_{ij}, \min_{i \ni G_q,\, j \ni G_k} d_{ij}\} \\
&= \min\{D_{pk}, D_{qk}\}
\end{aligned}
$$

由于 G_p 和 G_q 已并成一类，须将 $D(0)$ 的第 p、q 行及第 p、q 列删去(不妨设 $p<q$)。在第 p 行、q 列的位置记上 $D_{rk}(k=1,2,\cdots,m;k\neq p、q)$，形成新矩阵，并记作 $D(1)$。

(4) 对 $D(1)$ 重复对于 $D(0)$ 的步骤(2)，得 $D(2)$，由 $D(2)$ 按同样步骤，得 $D(3)$，…，直到所有的数据点都得到归类为止。如果某一步 $D(r)$ 中的最小元素不止一个，则对应的这些类可以同时合并。最后得到的聚类结果如图5-46所示。

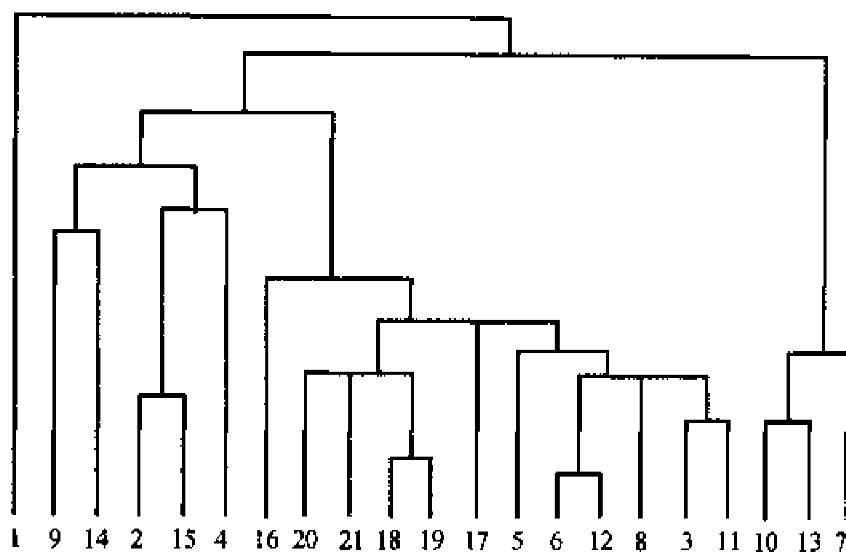


图5-46 聚类图

§6 空间数据的集合分析和查询

空间数据的集合分析和查询是指按照给定的条件，从空间数据库中检索满足条件的数据，以回答用户提出的问题，又称为咨询式分析。

通常，GIS的属性数据分为数字型和字符型两种形式，前者包括数量、等级等，用以表述地理实体的定量特征；后者包括名称、分类等，用以表述地理实体的定性特征。对于字符型数据，一般采用逻辑关系进行运算。对于数字型数据，则

可以进行"加"、"减"、"乘"、"除"、"乘方"等数学运算,以产生新的属性值。

6.1 空间集合分析

空间集合分析是按照两个逻辑子集给定的条件进行逻辑运算,其基本原理是布尔代数,它的运算符号或算子包括 *AND*、*OR*、*XOR*、*NOT* 及其组合等,逻辑运算的结果为"真"或"假"。

如图 5-47 所示,设属性分别为 *A* 和 *B* 的两个多边形相交,形成的空间单元分别为 1、2、3、4,则它们经布尔操作后的结果,如表 5-15所示。

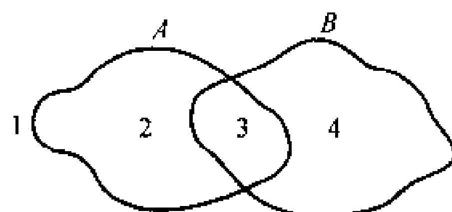


图 5-47 多边形相交及其形成的空间单元

表 5-15 逻辑运算的真(1)或假(0)

运算条件	运算结果			
	1	2	3	4
A·*OR*·(*NOTB*)	1	1	1	0
NOT NULL	1	1	1	1
NOT B	1	1	0	0
(*NOT A*)·*OR*·(*NOTB*)或 *A*·*XOR*·*B*	1	1	0	1
(*A*·*AND*·*B*)·*OR*·((*NOTA*)·*AND*·(*NOTB*))	1	0	1	0
(*NOTA*)·*OR*·*B*	1	0	1	1
(*NOT A*)·*AND*·(*NOTB*)	1	0	0	0
NOT A	1	0	0	1
A	0	1	1	0
A·*OR*·*B*	0	1	1	1
A·*AND*·(*NOTB*)	0	1	0	0
(*A*·*AND*·(*NOTB*))·*OR*·((*NOTA*)·*AND*·*B*)	0	1	0	1
A·*AND*·*B*	0	0	1	0
B	0	0	1	1
NULL	0	0	0	0
(*NOTA*)·*AND*·*B*	0	0	0	1

空间集合分析虽然都可以在基于矢量的 GIS 和基于栅格的 GIS 中完成,但基于栅格的系统要容易和快捷得多。例如,当利用栅格的游程编码数据进行空间聚类分析时,根据预先设定的聚类条件,从栅格数据中将所有符合条件的区域输出在图上,不符合条件的区域为空白。按照关系运算表达式,可将空间聚类

写为

$$C_k(U)\{(A,P)\in U|(A,P)\text{满足 }E\}$$

式中：$U=f\{L_1,L_2,\cdots,L_n\}$。

$L_1,L_2,\cdots,L_n$ 表示同一地区一组空间数据集合，它们按照统一的数学基础存储在地理信息系统的数据库中。现设 L_1 表示土层厚度大于 50 cm 的土壤单元的栅格数据 (A_i,P_i)；L_2 表示土壤质地为重壤的土壤单元的栅格数据 (A_j,P_j)；L_3 表示高程小于 30 m 的数字高程模型的栅格数据 (A_k,P_k)。则 E 的设定形式及其运算的结果如图 5－48 所示。

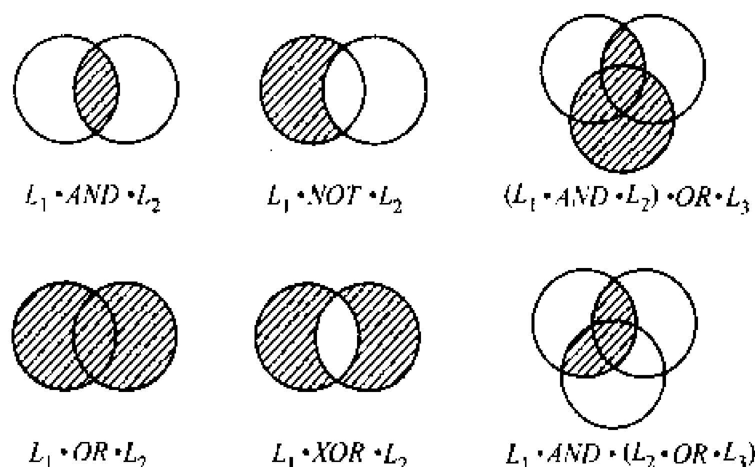


图 5－48 布尔逻辑算子及其运算结果的文氏图

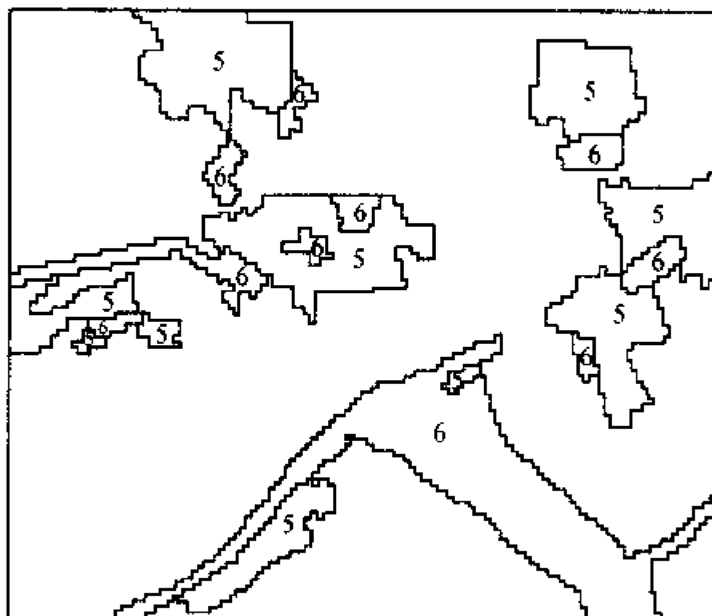


图 5－49 空间聚类分析输出
（本图的原始数据与图例说明参阅图 2－18）

根据图 5－48，当 $E=L_1\cdot AND\cdot L_2$ 时，将凡土层厚度大于 50 cm，且质地为重

壤的土壤单元输出在图上；当 $E = L_1\ OR\ L_2$ 时，将土层厚度大于 50 cm，或者质地为重壤的土壤单元都输出在图上，这时有些土壤单元可能同时满足两个条件，也可能只满足其中一个条件；当 $E = L_1\ XOR\ L_2$ 时，将同时不满足上述两个条件的土壤单元输出在图上；当 $E = L_1\ NOT\ L_2$ 时，将只能符合条件 L_1（其中质地为重壤的土壤单元除外）的土壤单元输出在图上。

E 可以按照需要的属性条件设定，也可以按照空间要素的几何条件设定。例如，E 的设定条件可以为

E =（属性 =“湖塘”∧面积≥525 m^2∧湖岸邻接居民地）

根据栅格数据输出的结果如图 5－49 所示。这种聚类条件的设定常用于位址规划。

6.2 空间数据的查询

空间数据的查询属于数据库的范畴，一般定义为从数据库中找出所有满足属性约束条件和空间约束条件的地理对象。属性约束条件一般用带比较运算符的逻辑表达式描述，这与传统的结构查询语言 SQL 的 where 语句中的条件表达式相似。空间约束条件用带空间谓词的逻辑表达式描述，空间谓词由地理对象间的空间关系演变而来，如包含、相交、分离、重叠、距离、方向等。因此，空间查询是作用在库体上的函数，返回用户请求的内容，也属于咨询式分析。

查询功能是 GIS 功能的重要组成部分，也是 GIS 面向用户最直接的窗口，GIS 用户提出的很多问题，通常都可以以查询的方式解决，但是不同的查询方法却在很大程度上制约着 GIS 功能的有效发挥。目前，GIS 中的空间查询大致可分为三类：① 针对空间关系的查询，例如，查询一条公路途经的所有城镇；② 针对非空间属性的查询，例如，查询一个城市的人口数量；③ 结合空间关系和非空间属性的查询，例如，查询距某条河流≥500 m、种植玉米且面积大于 53 hm^2 的土地利用单元。

空间查询的发展方向是开发强有力的空间查询语言，目前 GIS 的空间查询，主要有以下四种方法：

(1) 基于关系查询语言扩充的空间查询方法。由于空间数据库是以空间（地理）目标作为存储集，与一般数据库的最大不同点是它包含“空间”（或几何）概念，而标准的 SQL 是关系代数模型中的一些关系操作及其组合，适合于关系表的查询与操作，但却不支持空间概念和运算。因此，为支持空间数据的查询，需要在 SQL 上扩充谓词集，使之包含空间关系谓词，并增加一些空间操作。如 Egenhofer 根据空间数据库的特点以及空间数据表示的要求，在关系型的 SQL 上发展了一套空间结构化查询语言（Spatial SQL），在查询结果表示方面，提供了包含 6 个显示参数的显示环境。例如查询所有高速公路并用红虚线表示，用该空

间结构化查询语言表示为：

```
Set   Color Red
      Pattern Dashed
For   Select  Geometry
      From    Roads
      Where   Type = "Highway"
```

(2) 可视化空间查询方法。可视化空间查询是指将查询语言的元素，特别是空间关系，用直观的图形或符号表示。因为某些空间概念用二维图形表示比用一维文字语言描述清晰和易于理解，例如，某些空间操作可以图示如图5-50。

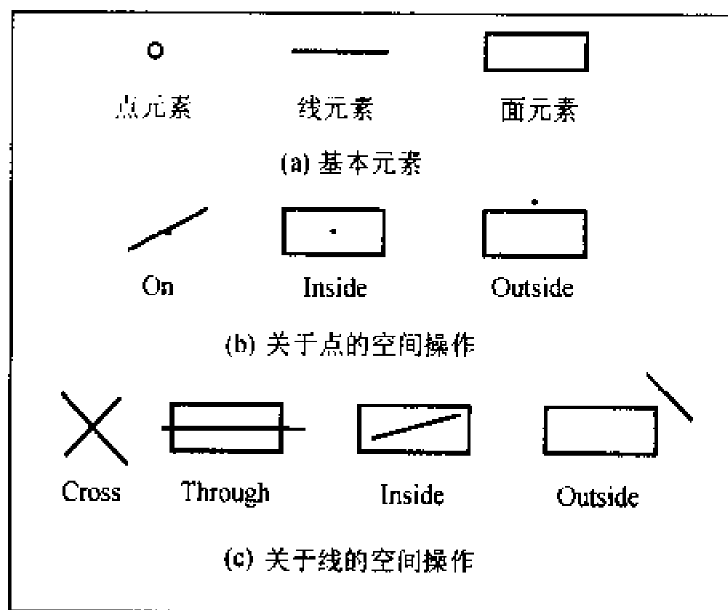


图5-50 可视化查询的语言元素

以图形、图像或符号为语言元素的可视查询，仅是对查询直观形象化的描述，它仍然要翻译成形式化的SQL语言。由于只有部分空间概念可用与人类空间观念一致或接近的图形或图像表示，某些查询仍然难以用人类易理解的图形表示，因此目前可视查询还不能表达所有的空间查询。

(3) 基于自然语言的查询方法。这是在查询语言中引人自然语言的概念，使空间查询变得更加简单和方便。例如，当查询高气温的城市时，可引人以下查询语言：

```
Select  Name
From    Cities
Where   Temperature is high
```

如果通过统计分析和计算，认为当城市气温大于或等于33.75 ℃时是高气温，则上述自然语言可转换为：

Selet Name

From Cities

Where Temperature > = 33.75

这里，在对自然语言中的模糊概念量化时，必须考虑当时的语言环境及背景。因此，引入自然语言的空间查询只能适用于某个专业领域的 GIS，而不能作为 GIS 中的通用数据库查询语言。

(4) 超文本查询方法。超文本(Hypertext)是由文本信息结点和表示文本信息结点间相关联的链所组成的一个具有一定逻辑结构和语义查询功能信息集成化网络。它的主要特征之一就是用户可以主动地决定阅读文本结点的顺序，不

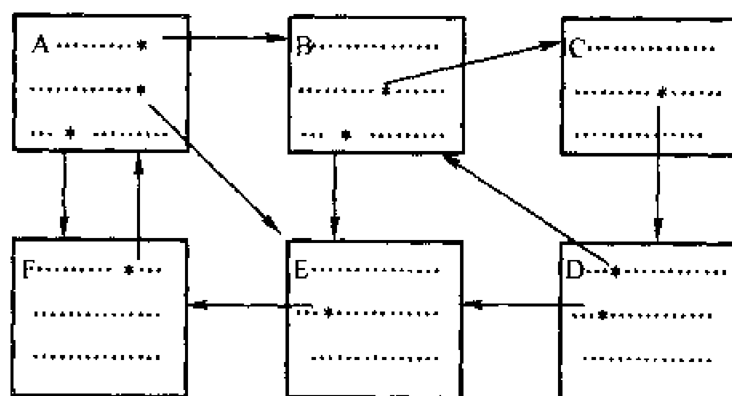


图 5-51 一个简单超文本结构示意图

(据王建华，1997)

像传统文本只能按照固定的线性顺序进行。如图 5-51 所示，超文本由若干内部关联的文本信息块组成，这些信息块是超文本的基本组成单元，它可以是图形、图像、字符组成的文本、文本库或帧等，统称其为超文本结点(node)，每个结点都有若干指向其他结点或从其他结点指向该结点的指针，称其为链(link)，每一条链连接着两个相关的结点，而且从一个结点指向另一个结点，这样的结点和链共同组成了超文本网络。当用户从标号为 A 的文本块开始查询时，则有三条路径可供选择，其一是从 A 到 B，其二是从 A 到 E，其三是从 A 到 F，如用户选择 A 到 B，则可继续选择 B 到 C 或 B 到 E 等。由于链是按用户的要求预先设计的，可以保证并实现超文本"联想"式的查询，但用户不能实时构建自己需要的各种查询。

思考题

1. 试解释缓冲区分析和叠合分析的概念，并举例说明这两种空间分析方法的用途。

2. 比较缓冲区查询与缓冲区分析的概念。

3. 设某一实体的影响度 F_i 随距离 r_i 而呈指数率变化，已知该实体的影响半径为 d_0，分级指标值为 f_0，试述对该实体进行缓冲区分析的步骤和方法($i=1,2,\cdots,5$)。

4. 讨论从数字高程模型(DEM)中提取河网过程中可能遇到的算法问题,请您提出解决问题的有效方法。

5. 写出根据矢量法计算地表微分单元坡向的公式,并绘图表示当坡向分别为南、东、北、西时的 $\vec{n}_{i,j}$ 及 θ 角的状态。

6. 如图 5－52,设两个输入数据层的 A、B、D 分别表示属性类别数据的代码,图中有阴影像元表示没有被正确分类的像元。试问,通过 GIS 的逻辑"交"("AND")和逻辑"并"("OR")运算后,得到的输出数据层的精度有何不同?

A	A	B	B	B
A	A	B	B	B
D	A	A	B	B
D	D	D	D	B
D	D	D	B	B

图层1

A	A	A	B	B
A	A	A	B	B
A	A	A	B	B
A	A	D	D	D
A	D	D	D	D

图层2

图 5－52

第6章 地理信息系统的应用模型

§1 GIS应用模型概述

1.1 GIS应用模型的分类

如前所述，由于地理信息系统具有优良的硬件环境，多功能的软件模块，能客观地表达地理空间的数据模型，以及便于沟通人机联系的用户界面，使系统具有广泛的用途。地理信息系统的应用模型，就是根据具体的应用目标和问题，借助于GIS自身的技术优势，使观念世界中形成的概念模型，具体化为信息世界中可操作的机理和过程。这种模型的构建，不但是解决实际复杂问题的必要途径，而且也是GIS取得经济和社会效益的重要保证。

许多实际复杂的客观问题及其过程，往往是由若干可以计数的重要要素所决定的，正如爱因斯坦曾经指出的“物理上真实的东西一定是逻辑上简单的东西，也就是说，它在基础上具有统一性。”[①] GIS应用模型的作用，正是用一定程度的简化和抽象，通过逻辑的演绎，去把握地理系统各要素之间的相互关系、本质特征及可视化显示。

GIS应用模型根据所表达的空间对象的不同，可将模型分为三类(表6-1)：一类是基于理化原理的理论模型，又称为数学模型，是应用数学分析方法建立的数学表达式，反映地理过程本质的理化规律、如地表径流模型、海洋和大气环流模型等；一类是基于变量之间的统计关系或启发式关系的模型，这类模型统称为经验模型，是通过数理统计方法和大量观测实验建立的模型，如水土流失模型、适宜性分析模型等；还有一类是基于原理和经验的混合模型，这类模型中既有基于理论原理的确定性变量，也有应用经验加以确定的不确定性变量，如资源分配模型、位置选择模型等。

按照研究对象的瞬时状态和发展过程，可将模型分为静态、半静态和动态三类。静态模型用于分析地理现象及要素相互作用的格局；半静态模型用于评价

① 引自《爱因斯坦文集》，第一卷，人民出版社，第380页。

应用目标的变化影响;动态模型用于预测研究目标的时空动态演变及趋势。

表6-1 地学模型分类

模型分类	理论依据	应用领域	模型
理论	物理或化学原理	地表径流	运动方程
混合	半经验性	资源分配	运输方程
经验	启发式或统计关系	水土流失	统计、回归

注:据陈述彭等,2000。

目前,GIS技术的应用,已经从数据存储管理和查询检索,演进到以时空分析为主体,正在向着支持区域系统空间结构演化的预测、动态模拟及其空间格局的优化的发展新阶段。科学预测、动态模拟和辅助决策是GIS应用的高层次阶段,构建区域空间动力学应用模型将是区域可持续发展研究和GIS应用向纵深发展的交汇点。

1.2 GIS应用模型的构建

用GIS求解问题的过程如图1-13所示,实际包括目的导向(goal-driven)分析和数据导向(data-driven)操作两个过程。目的导向分析,是将要解决的问题与专业知识相结合,从问题开始,一步步地推导出解决问题所需要的原始数据、精度标准、模型的逻辑结构和方法步骤。数据导向操作,是将已经形成的模型逻辑结构与GIS技术相结合,从各类数据开始,一步步地将数据转换为问题的答案,必要时还需要进行反馈和修改,直到取得满意的结果,最后以图形或图表的形式输出最终结果。

GIS应用模型的构建,通常采用以下三种不同的途径:

(1) GIS环境内的模型建造。指应用者利用GIS软件的宏语言(如ARC/INFO的AML和System9的ATP等)发展各自所需的空间分析模型。这种模型法是将由GIS软件支持的功能看作模型部件,按照分析目的和标准,对部件进行有机的组合。因此,这种构模法能充分地利用GIS软件本身所具有的资源,模型建造和开发的效率比较高。

(2) GIS外部的模型建造。这种方法是基于应用GIS的空间数据库和输出功能,而模型分析功能则主要是利用其他应用领域的软件。这种构模法虽然运行效率受到很大影响,但实现了软件的嫁接,无需在GIS环境中重编分析软件,并具有广泛的适用性。

(3) 混合型的模型建造。这是上述两种建模法的结合,即尽可能利用GIS提供的功能,最大限度地减少用户自行开发的压力,又不失具有外部建模法的灵

活效果,例如自然语言命令或地图代数法,就属于这类方法。

这类方法常用的自然语言命令,例如:

① Average Exposure Times 2 Plus Steepness Times 3 For Cost Analysis(益本分析);

② Subtract V Minus U For Change Analysis(动态分析);

③ Renumber Landuse For Water Assigning 0 To 1 ~ 3,5 ~ 6,And 1 To 4(聚合分析);

④ Add Overlay 1 To Overlay 2 For Overlay3(叠加分析);

⑤ Spread Roads To X Meters(缓冲区分析)。

为了了解以计算机为中心的地理信息系统技术,如何解决地学和工程领域里的实际问题,包括资源的调查与评估、土地潜力与适宜性分析、人口预测与位置选择、环境监测与预报,以及在城市与交通的规划和管理中的应用等,以下列举一些具体的应用实例,既提供地理信息系统应用的一些数学模型,也展示一下GIS与某些产品类型相结合的应用前景与潜力。

§2 适宜性分析模型

适宜性分析是指土地针对某种特定开发活动的分析,这些开发活动包括农业应用、城市化选址、作物类型布局、道路选线、选择重新造林的最适宜的土地等。因此,建立适宜性分析模型,首先确定具体的开发活动,其次选择其影响因子,然后评判某一地域的各个因子对这种开发活动的适宜程度,以作为土地利用规划决策的依据。过去这种适宜性分析一般是采用各个因子简单叠合分析或通过地图覆盖的方法来解决,20世纪60年代后期,由于重视开发活动引起的环境效应,以及系统论方法在土地利用规划中的应用,逐渐利用以计算机为中心的地理信息系统进行土地利用规划的研究。

2.1 一般形式

设有某项评价目标或开发活动 T,该 T 对应一组影响因素 $X_1, X_2, \cdots, X_m$;每个因素对应一组参评因子 $x_1, x_2, \cdots, x_l$;每个因子有一组属性 $v_1, v_2, \cdots, v_m$;因此,每个因素对应一个属性集 V_i:

$$V_i = [v_{11}, \cdots, v_{jk}, \cdots, v_{nl}]$$

$$i = 1,2,\cdots,m; j = 1,2,\cdots,n; k = 1,2,\cdots,l。$$

显然,每个因素的属性集都是一个对指定的 T 从优到劣的全序集,且满足:

$$V_{1l} > V_{jl} > \cdots > V_{nl}$$

各个参评因子及其属性值的取得由数据库提取或由GIS空间分析软件生成。各个因素按其属性集的优劣，可用下列矩阵表示：

$$R = \begin{bmatrix} W_1 P_{11} & \cdots & W_i P_{1i} & \cdots & W_m P_{1m} \\ \vdots & & \vdots & & \vdots \\ W_1 P_{j1} & \cdots & W_i P_{ji} & \cdots & W_m P_{jm} \\ \vdots & & \vdots & & \vdots \\ W_1 P_{n1} & \cdots & W_i P_{ni} & \cdots & W_m P_{nm} \end{bmatrix} \quad (6-1)$$

式中：P 为 X_i 对 T 的贡献函数值；

W 为 X_i 对 T 的权重值。

P 值的确定方法为：将各因子最适宜的指标值定为贡献函数值100，将各因子最不适宜的指标值定为贡献函数值0，在这之间，指标值与函数值按线性关系计算和确定。

有了上述矩阵数据和GIS功能的支持，可以求取基于栅格单元的评价分值：

$$R(T) = \frac{1}{100}\sum_{i=1}^{m} W_i P_{ji}, j = 1,2,\cdots,n \quad (6-2)$$

然后，根据使

$$G(P_{ji}) = 1 - | R(T) - P_{ji}/100 | \quad (6-3)$$

的值为最大时的 P 所对应的 j，即为所求的某个 T 的适宜级 S_j。显然，当适宜级为 S_2 或 S_3 时，必须同时确定其限制性因子。限制性因子的计算公式为

$$L(x_i) = f \max_{1 \leq i \leq m} [| 100 - P(x_i) | W_i]。 \quad (6-4)$$

2.2 应用实例

地理信息系统技术在土地评价中的应用，已有许多成功的范例。这里以某地区的玉米种植用地评价为例，介绍适宜性分析模型的应用及基于GIS的土地评价方法和过程。

(1) 评价对象。玉米种植用地的土地适宜性评价，通过评价将研究区分出不同的适宜性等级：S_1(最适宜)、S_2(次适宜)、S_3(临界适宜)和 N(不适宜)。

(2) 评价方法。采用基于GIS的土地质量评价法，即将玉米作物生长有关主导生态条件与土地质量(供水、供肥等)相比照，从而评定土地的适宜性等级。

(3) 评价过程。如下。

① 评价对象生态条件的调查。评价对象玉米属于禾本科，为一年生草本，

其主导生态条件例如：性喜高温，需水量大，要求土壤肥沃和土层疏松，其根系生长要求防止土壤侵蚀等。

② 确定评价对象的影响因素和因子。根据将玉米作物生长有关的生态条件与土地质量相比照，除了温度可通过季节调节外，其他影响因素和因子如图6-1所示：

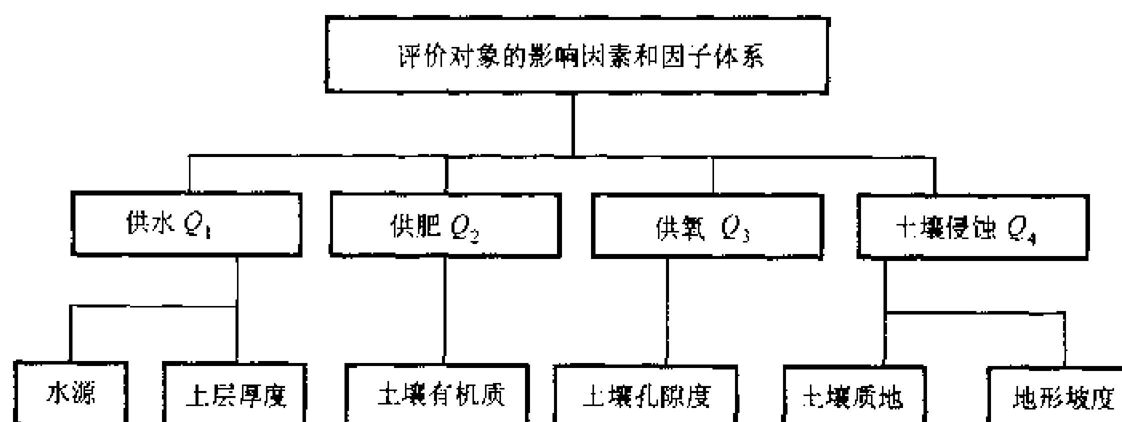


图6-1　玉米作物生长有关的影响因素和因子分析图

③ 利用GIS生成影响因素数据。例如供水有效性，其影响因子为水源和土层厚度，它们对供水或保水有效性的属性集如表6-2所示：

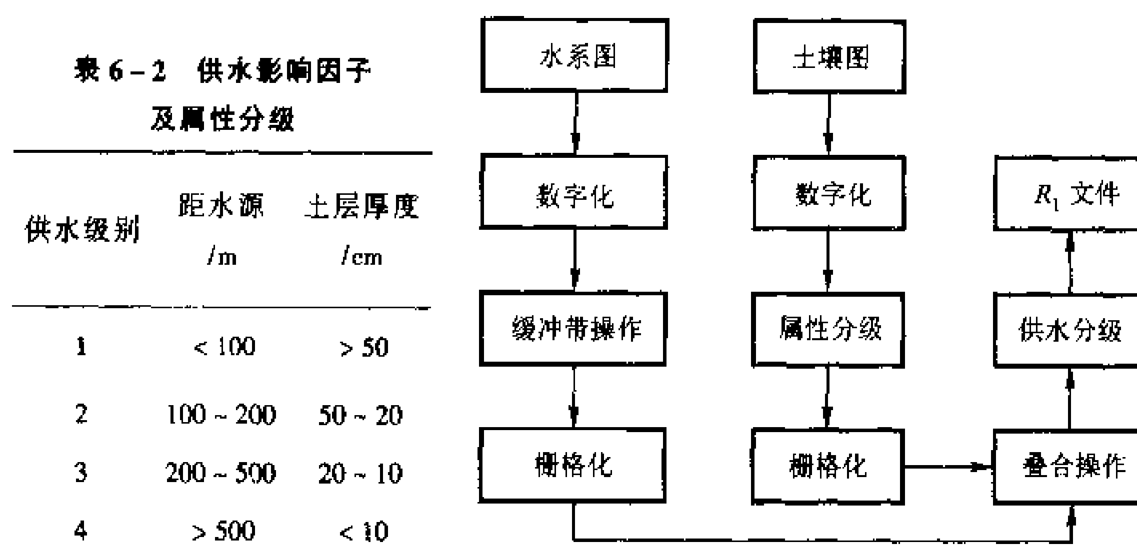


表6-2　供水影响因子及属性分级

供水级别	距水源 /m	土层厚度 /cm
1	< 100	> 50
2	100～200	50～20
3	200～500	20～10
4	> 500	< 10

图6-2　GIS生成供水因素数据的操作过程

根据表6-2和相应的空间数据，通过如图6-2所示的GIS操作过程，可以有效生成基于栅格单元的供水条件等级数据文件 R_1。同理，可以生成供肥条件等级数据文件 R_2、供氧条件等级数据文件 R_3 和土壤侵蚀等级数据文件 R_4。这些数据文件是土地适宜性评价的重要基础，而它们的可靠性又取决于各个参评

因子及其属性集的确定。

④ 计算各因素的权重和贡献函数值。由于每个影响因素或土地质量 $Q_i(i=1,2,3,4)$对评价对象的贡献大小和适宜程度不同,因此不同因素有不同的权重 W_i,不同级别的因素对评价对象的贡献函数值 P_{ji}也不相同。

权重 W_i 的计算方法为:首先,将土地质量按其对评价对象重要性的不同,进行由小到大的排队;其次,确定 Q_{i+1}对 Q_i 的重要性程度 $V_{i+1,i}$,用倍数表示,并令 $V_{1,0}=1$,其结果如表 6-3 所示;最后,按照下列公式计算各土地质量的权重:

表 6-3 土地质量重要性比较表

土地质量 Q_i	土地质量重要性比较及取值 $V_{i+1,i}$
$Q_4(Q'_1)$	$1(V_{1,0})$
$Q_3(Q'_2)$	$1(V_{2,1})$
$Q_2(Q'_3)$	$2(V_{3,2})$
$Q_1(Q'_4)$	$4(V_{4,3})$

$$W_i = U_i / \sum_{i=1}^{m} U_i$$

式中,$U_1=V_{1,0}$,$U_2=V_{2,1}\times V_{1,0}$,$\cdots U_i=V_{i,i-1}\times V_{i-1,i-2}\times\cdots\times V_{1,0}$,所以,和土地质量 $Q_i(i=1,2,3,4)$对应的权重分别为 $W_1=0.67$,$W_2=0.17$,$W_3=0.08$,$W_4=0.08$。

贡献函数值可以按照影响因素的级别来确定。例如,如果将各因子最适宜的指标值定为贡献函数值 100,将各因子最不适宜的指标值定为贡献函数值 0,当影响因素分为 4 级时,则各因素的 P 值分别为

$$P_{1i}=100, P_{2i}=67, P_{3i}=33, P_{4i}=0$$

根据各因素的权重与贡献函数值的关系式,可以建立各因素的评价指标表(表 6-4),该指标表与不同土地质量的数据文件 R_i 相结合,为土地适宜性评价提供依据。

⑤ 计算机适宜性评级。通过将不同土地质量数据文件 R_i 的等级,切换为与表 6-4 相同等级对应的指标值,便可计算出基于栅格单元的评价分值 $R(T)$:

$$R(T) = \frac{1}{100}\sum_{i=1}^{m} W_i P_{ji}, j = 1,2,\cdots,n$$

表 6-4 因素、等级评价指标表

指标 因素 等级	Q_1	Q_2	Q_3	Q_4
1	67.0	17.0	8.0	8.0
2	44.8	11.4	5.4	5.4
3	22.2	5.6	2.6	2.6
4	0	0	0	0

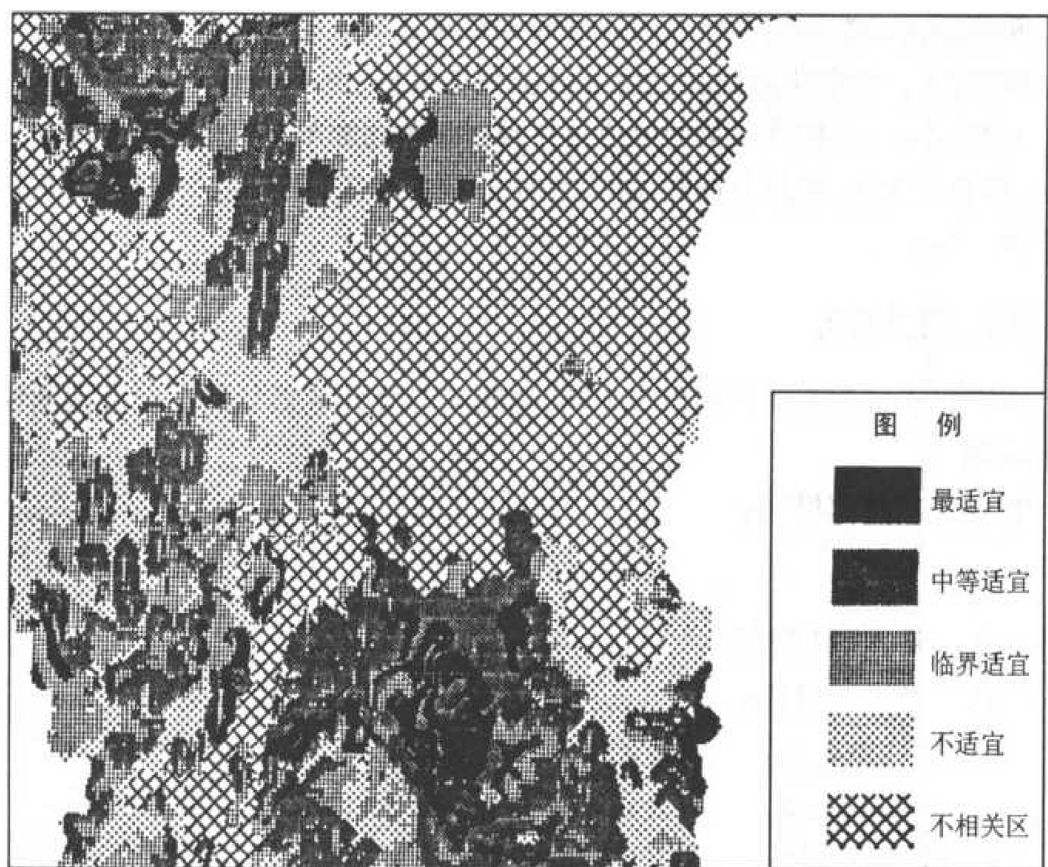


图 6-3 玉米作物种植用地适宜性分级图

然后，根据使 $G(P_{ji}) = 1 - | R(T) - P_{ji}/100 |$
的值为最大时的 P 所对应的 j，即为所求的对 T 的适宜性等级 S_j。例如，当某栅格单元 Q_1 的指标值为 67.0，Q_2 为 11.4，Q_3 为 2.6，Q_4 为 0 时，该栅格单元的 $R(T)=81.0$，然后将该值和 P_{ji} 的值依次代入式(6-3)进行计算，只有当取 $P_{2i}=67$ 时，$G(P_{2i})$的值为最大(0.86)，所以该栅格单元的适宜性等级为 S_2(次适宜)。直至研究地区全部栅格单元都获得相应的适宜性等级，得到玉米作物种植

用地的适宜性分级图(图6-3)。

§3　发展预测模型

3.1　一般介绍

发展预测是运用已有的存储数据和系统提供的手段，对事物进行科学的数量分析，探索某一事物在今后的可能发展趋势，并作出评价和估计，以调节、控制计划或行动。在地理信息研究中，如人口预测、资源预测、粮食产量预测以及社会经济发展预测等，都是经常要解决的问题。

预测方法通常分为定性、定量、定时和概率预测。在信息系统中，一般采用定量预测方法，它利用系统存储的多目标统计数据，由一个或几个变量的值，来预测或控制另一个研究变量的取值。这种数量预测常用的数学方法有移动平均数法、指数平滑法、趋势分析法、时间序列分析法、回归分析法，以及灰色系统理论模型的应用。

3.2　应用实例

以下以人口和劳动力的预测为例，说明人口统计数据如何在定量预测模型中的应用。

根据人口预测模型：

$$P_t = P_0 e^{(\lambda - u)}$$

式中：P_t　第 t 年人口数；

P_0　基年人口数；

λ　人口出生率；

u　人口死亡率；

t　时间(年份)。

设根据研究地区一组人口统计数据的分析，得 $\lambda \approx 12.5\%$，$u = 6.5‰$，将基年定为1985年，并且 $P_0 = 612.7$ 万人。设每年净迁入该研究地区的人口数为 $W = 5$ 万人。则

$$\begin{aligned} P_1 &= P_0 e^{(\lambda - u)} + W \\ P_2 &= P_1 e^{(\lambda - u)} + W \\ &\vdots \qquad \vdots \quad \vdots \\ P_t &= P_{t-1} e^{(\lambda - u)} + W \end{aligned}$$

于是可得到规划期的人口预测数(表6-5)。

表 6-5 规划期人口预测数

年份	1985	1986	1987	1988	1989	1990	增长速度
人口数/万人	612.7	621.4	630.1	638.9	647.7	656.6	1.43%

同理,根据劳动力预测方程:

$$L_{(t)} = LR_{(t)} \cdot L_{(t-1)} + LW_{(t)}$$

式中:$L_{(t)}$为第 t 年劳动力状态向量,即 $L_{(t)} = (L'_{18}, L'_{19}, \cdots, L'_{60})$;

$LW_{(t)}$为第 t 年劳动力迁移向量,即 $LW_{(t)} = (LW'_{18}, LW'_{19}, \cdots, LW'_{60})$

$LR_{(t)}$为劳动力存留系数矩阵,即

$$LR_{(t)} = \begin{bmatrix} r'_{18} & & & 0 \\ & r'_{19} & & \\ & & \ddots & \\ 0 & & & r'_{60} \end{bmatrix}$$

式中下标 18~60 表示劳动力的年龄;

r 代表分年龄层的劳动力存留比率。

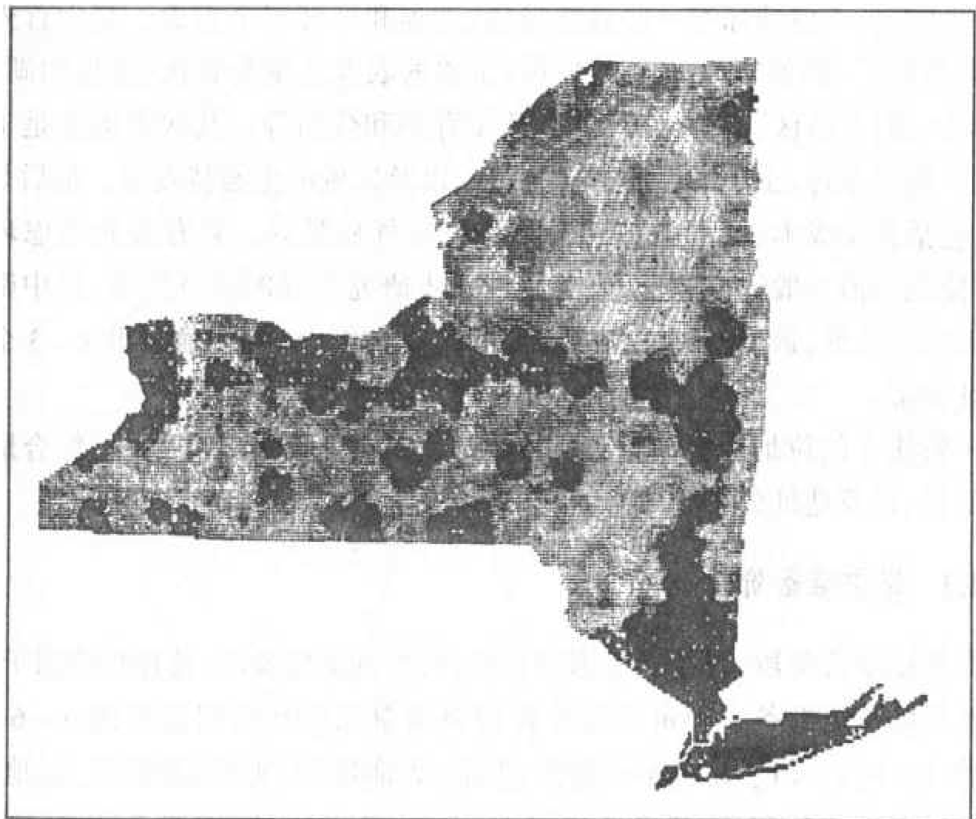


图 6-4 美国纽约州 1990 年人口发展预测图

(据 H.W.Calkins 等,1977)

于是，得到研究地区规划期劳动力的预测数(表 6－6)。

表 6－6 规划期劳动力预测数

年份	1985	1986	1987	1988	1989	1990	增长速度
劳动力人数/万人	335.1	341.4	348.3	353.4	359.5	367.1	2.17%

有了这些预测的结果，将这些结果与表示每个镇、市中心点的 x、y 坐标联系起来，便得到一组点的数据，这组数据加上研究地区的边界数据，输入 SYMAP 程序，通过使用绘制等值线这个权，便输出一幅人口发展预测图(图 6－4)，该图表示出美国纽约州 1990 年的人口密度，原始输出图经过缩小，使图形效果更为直观，概括地显示出所预测的人口的增长趋势，作为区域经济发展规划的依据，以便寻找对策，使人口的增长与有效的土地面积和其他的资源相适应。

§4 位址选择模型

位址选择是指按照规定的标准，通过空间分析的方法，确定厂址、电站、管线，或者交通路线等的最佳位址或路径。

位址选择考虑的标准一般包括环境、工程和经济三个方面。其中首先考虑的是环境标准，例如 20%以上的坡度，主要的农业土壤分布区、湿地和湖区、文化活动区、国有林区、资源保护区，以及体育场和公园等。其次考虑的是工程标准，包括地形条件，土壤的性质，气候因素，以及区域的生态特点等。最后是经济标准，包括开发成本、供水条件、铁路运输、空气质量等。只有首先考虑环境标准，才能识别出一般适合的位置，然后进一步研究工程的经济因素，从中筛选出优先考虑的位址，最后通过详细的环境和工程的综合论证，确定出 1～3 个最佳的选址方案。

一般建立的位址选择模型如图 6－5 所示，分为数据准备阶段、综合影响的评价阶段，以及选址分析阶段。

4.1 数据准备阶段

在数据准备阶段，要建立专家咨询组，明确选址的要求，选择影响因子，进行位址选择的数据准备。确定位址分析的格网单元的数据记录如图 6－6 所示。图中的 $V_1, V_2, \cdots, V_m$ 分别表示高程、土壤、表面地质、水域、易泛区、湿地、森林覆盖、开阔地、市镇用地、居民地规划区、农田、公共土地、资源保护区、道路、土地利用等数据变量。根据土地利用数据，首先进行位址的初步甄别，将环境的敏感区域和保护区域，从位址选择的空间范围中排除出去。然后利用地形图和

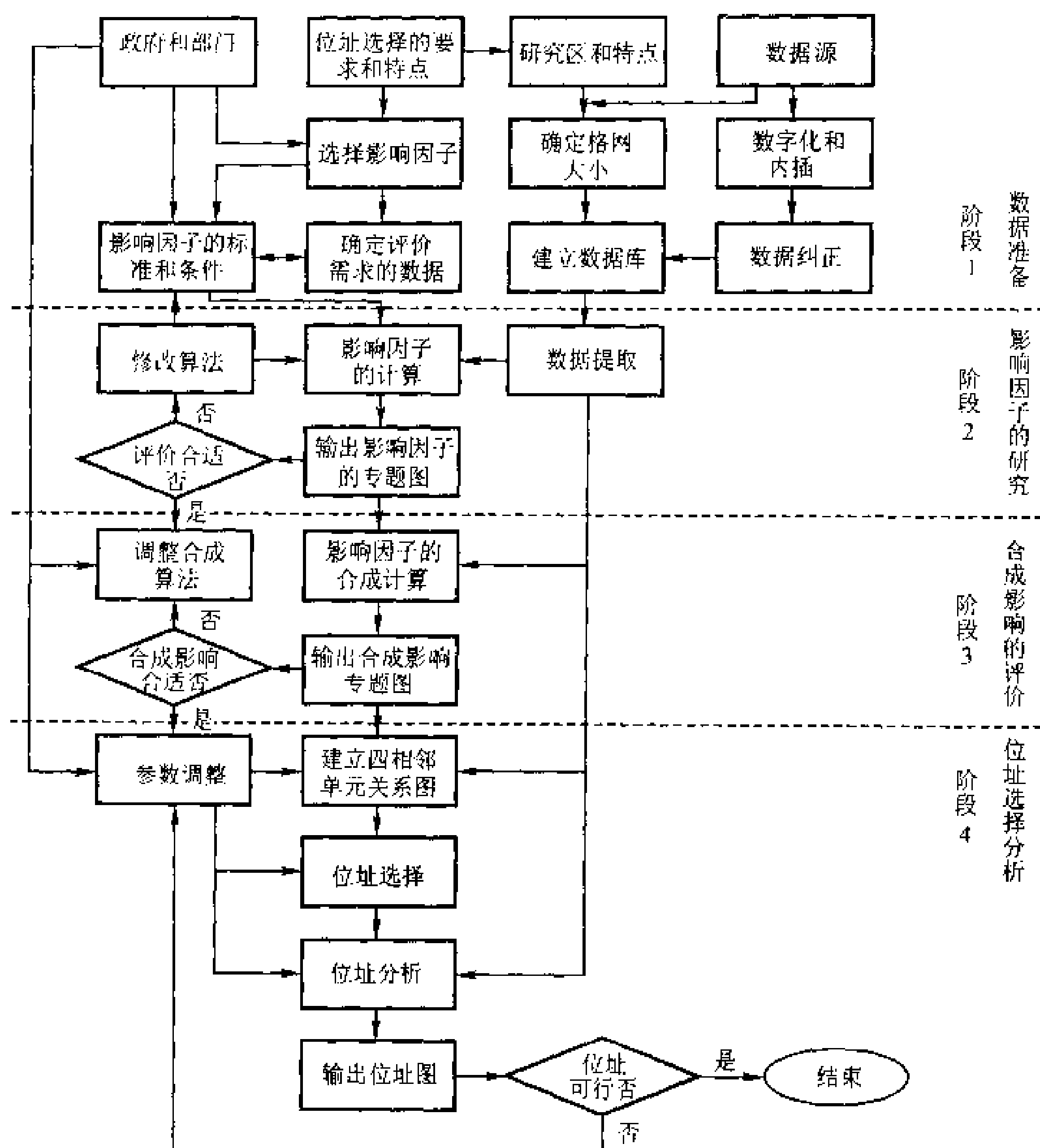


图6-5　位址选择模型

存储的数据，分别计算影响因子的各个因素向量，各个因素向量按六级分类法，分为无影响、轻微影响、中等影响、重要影响、严重影响和禁止开发。最后输出各个影响因子的分析图，同时建立影响因子的格网数据文件，提供综合评价和分析。

4.2　综合影响评价阶段

综合评价阶段的任务是按照工程和经济可行性的要求，建立选址条件、综合评价的标准和算法，例如，

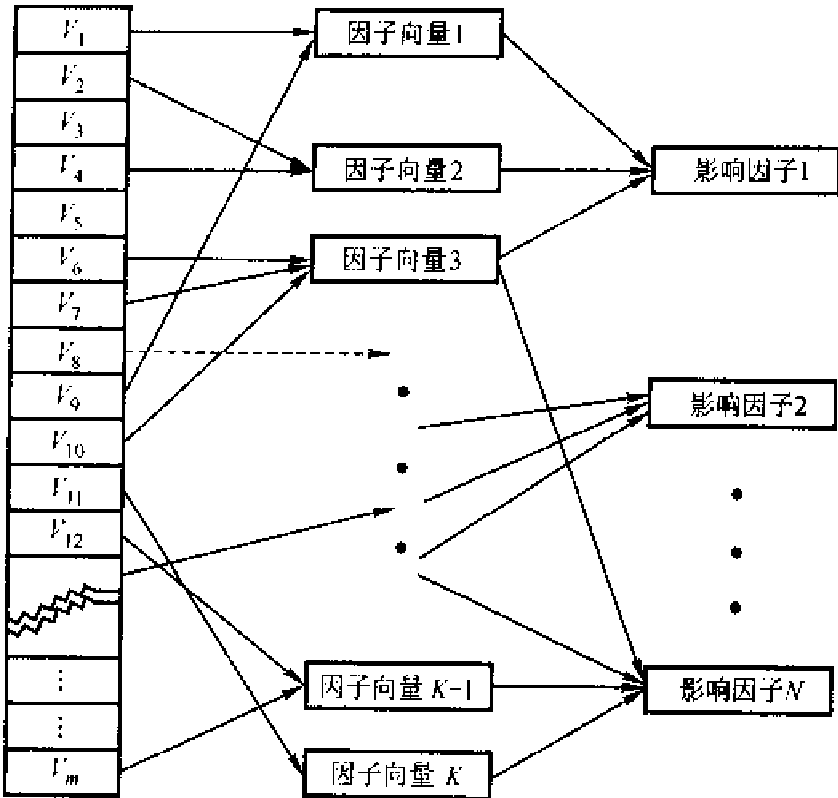


图6-6 位址选择数据准备

$$条件 = (S \wedge C \wedge F \wedge L \wedge E)$$

式中：S 为坡度<5%；

C 为开发成本适宜；

F 为离开居民区；

L 为地耐力坚固；

E 为环境质量优良。

于是，根据各个影响因子可以进行综合影响的评价(图6-7)。

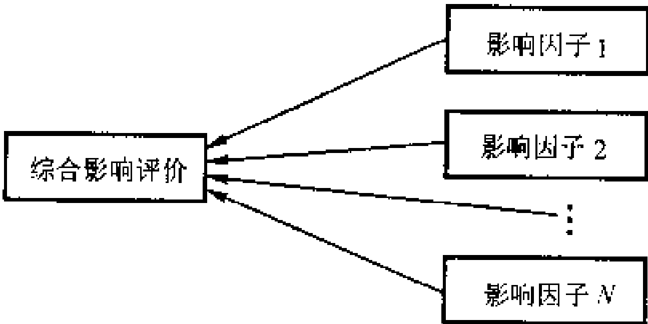


图6-7 综合影响评价

开发成本的计算根据区域单元的影响因子(表6-7)和选址路线的距离

(图 6-8)进行,例如从 CA 至 CB 的成本,可按下式计算:

$$成本 = DA \cdot A\ 区因子 + DB \cdot B\ 区因子$$

表 6-7 按影响级别确定的单位距离成本

影响级别	成本		
	A 区	B 区	C 区
0	0	0	0
1	1	1	1
2	2	2	2
3	4	4	4
4	8	8	6
5	16	16	256
6	32	∞	∞

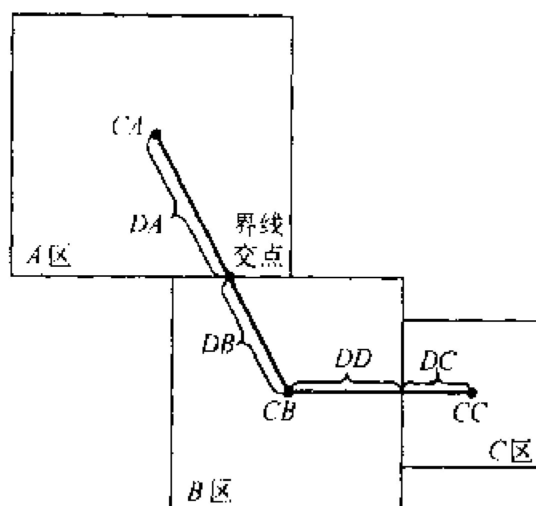


图 6-8 选址路线在格网中通过的距离

4.3 位址选择分析阶段

位址选择分析阶段的任务是实施位址的选择,并对结果进行分析评价。其运行过程如图 6-9 所示。

实施位址选择是根据综合影响的评价数据(图 6-10),即首先根据该评价数据建立相邻单元关系图(图 6-11),由相邻单元关系图,通过交点之间数据的比较,建立顶点连接图(图 6-12);由顶点连接图可以算出各条边的权值,如表 6-8所示;由该表可以建立对应于图 6-12 的权矩阵 W:

$$W = [W_{ij}] \tag{6-5}$$

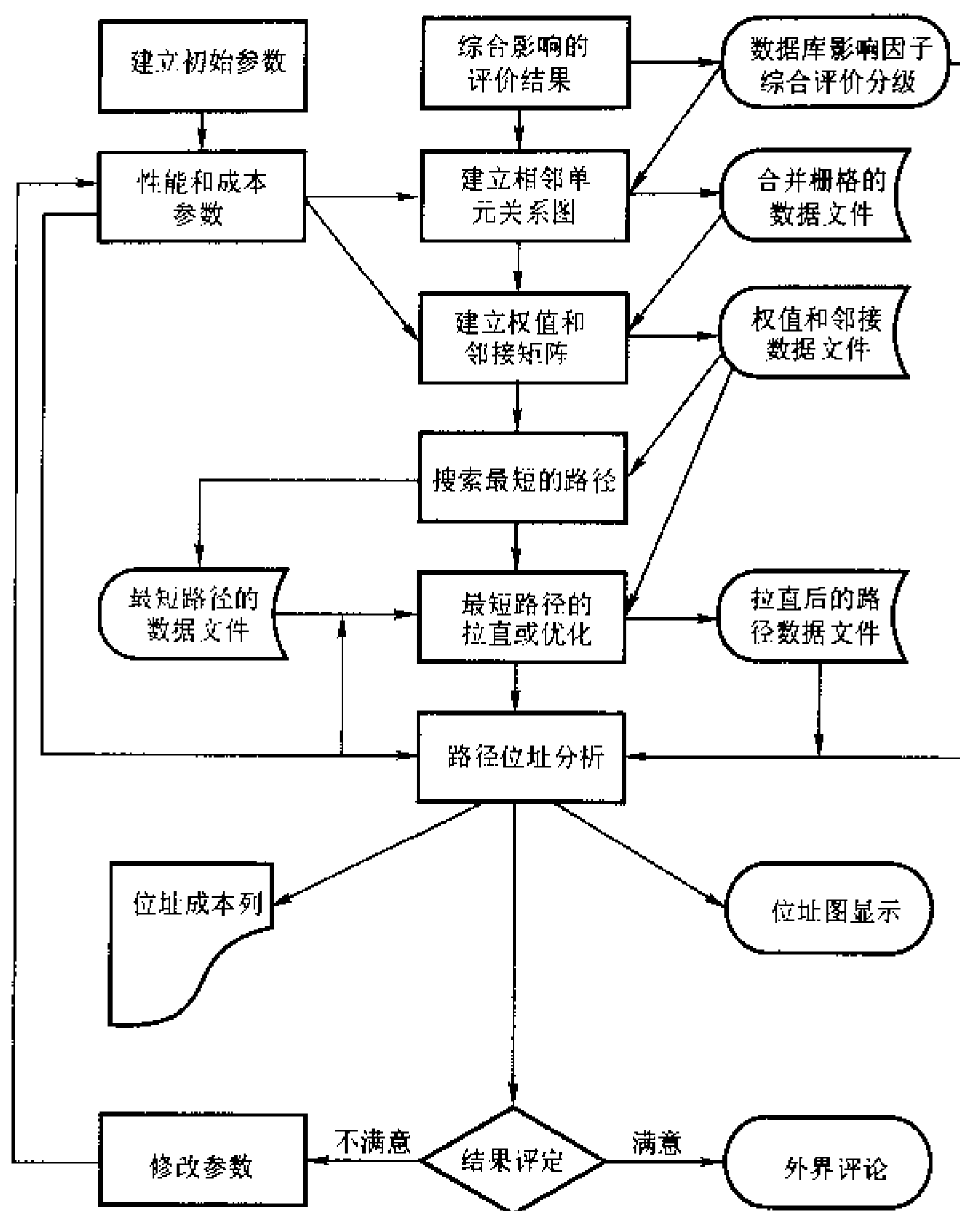


图 6-9　位址选择分析阶段的运行过程

式中，元素 W_{ij} 为图中的边 e_{ij} 的权，并规定当 i,j 间有边连接时，$W_{ij}>0$；当 i,j 间无边连接时，$W_{ij}=\infty$；当 $i=j$ 时，$W_{ij}=0$；对于无向图，$W_{ij}=W_{ji}(i\neq j)$。

1	1	5	1
1	1	6	4
4	3	3	3
1	3	3	3
2	3	3	3

图 6-10　综合评价数据

⇒

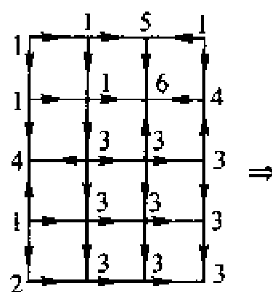


图 6-11　相邻单元关系图

⇒

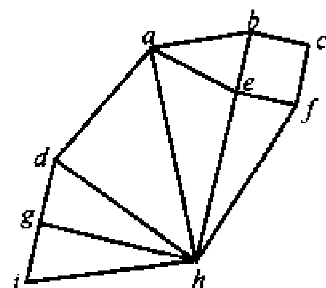
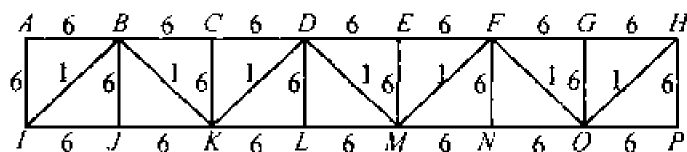


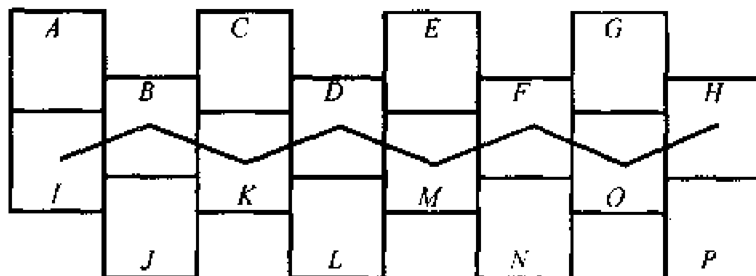
图 6-12　顶点连接图(无向图)

表 6-8 图 6-12 的各边权值

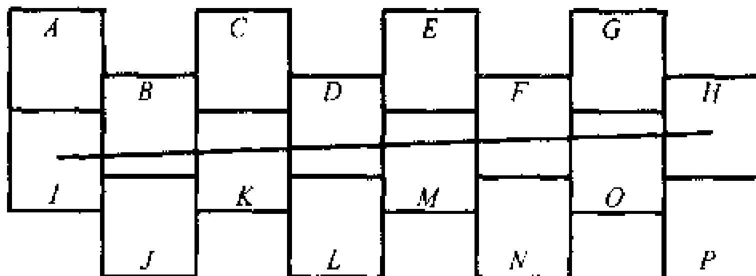
边	权	边	权
(a,b)	1.97	(e,h)	2.00
(a,e)	1.97	(f,h)	2.23
(a,d)	1.97	(d,h)	2.23
(a,h)	2.92	(d,g)	1.00
(b,c)	1.00	(g,h)	2.00
(b,e)	1.00	(g,i)	1.00
(c,f)	1.00	(h,i)	2.23
(e,f)	1.00		



(a) 最低成本路径：I、B、K、D、M、F、O、P



(b) 区域单元的连接



(c) 路径的裁直

图 6-13 路径变换处理示例

图 6-12 的权矩阵 W 如式 6-5 所示。于是，输入该权矩阵，并设待确定的最短路径的起始节点为 K，终止节点为 L，根据戴克斯特拉（Dijkstra）提出的算法，可以进行最短路径的设计（见第 5 章 §4）。设根据该算法得到最低成本的路径 $I—B—K—D—M—F—O—H$（图 6-13(a)），通过区域单元的连接，得到路径的位址（图 6-13(b)和图 6-14），再经过裁直处理，得到优化后的路径位址（图6-13(c)和图 6-15）。

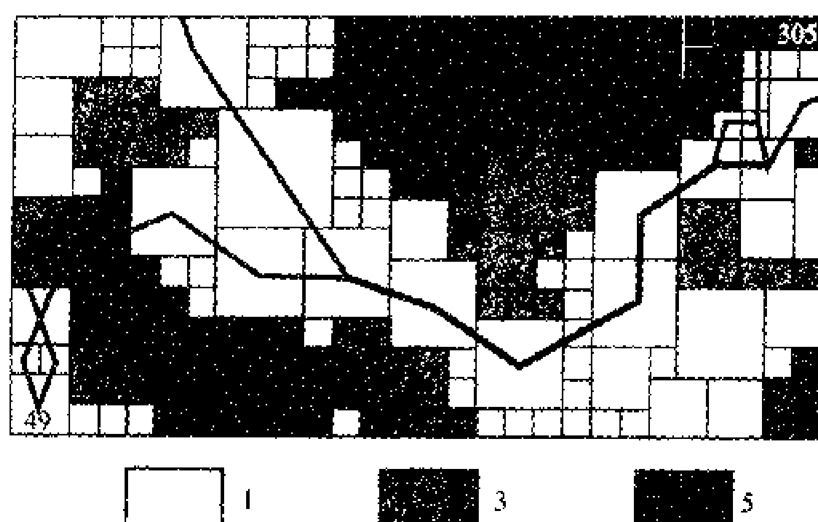


图6-14　路径连接处理示例

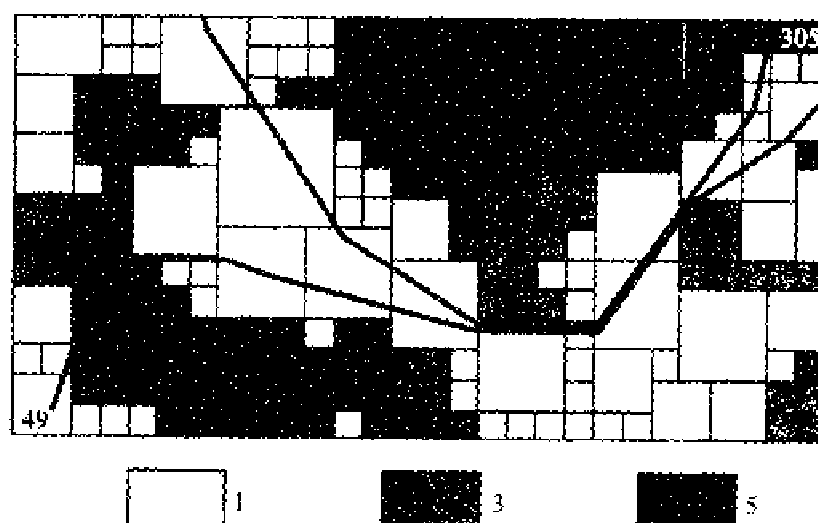


图6-15　路径裁直处理示例

§5　交通规划模型

城市交通模型包括城市交通发生量预测，出行分布预测和交通量最优分配三个部分。目前，普遍采用的是 O～D 调查的方式了解城市交通的发生、分布和规划分析。

5.1　交通发生量预测模型

该模型采用因果分析法，综合考虑影响交通量发生的各因素，用回归分析法建造多因素相关回归方程。

设有 n 个居民小区，m 种出行目的，L 个影响因素。

设 r_{ij} 为区域 i 居民第 j 类出行目的的出行发生率，X_{ij}^k 是区域 i 居民第 j 类出行目的的第 k 个影响因素。

则建构回归方程

$$r_{ij} = a_0 + \sum_k a_k X_{ij}^k$$

$$i = 1,2,\cdots,n$$

$$j = 1,2,\cdots,m$$

$$k = 1,2,\cdots,L$$

以及出行交通发生量 T 为 $T = \sum_i \sum_j r_{ij}$

其中，第 j 种出行目的的交通发生量 $T_j = \sum_i r_{ij}$

i 小区的交通发生量即为 $T_i = \sum_j r_{ij}$

式中：a_0, a_k 由回归分析确定。

j 可以取平均每户就业率、到出行目的地的距离、家庭收入水平等等。

5.2 出行分布预测模型

出行分布包括出行方向、出行数量以及出行工具的空间分配，我们主要考虑以居民区为出发点的出行分布情况。

设 T_{ij}^k 为从 i 居民区到 j 出行目的地用 k 种交通工具的交通量。

则一般的出行分布预测模型为 $T_{ij}^k = A_i B_j O_i D_j f(C_{ij}^k)$

式中：A_i 为与 i 居住区有关的比例因子；

B_j 为与 j 区域有关的比例因子；

O_i 为 i 居住区的总出行量；

D_j 为 j 区域的吸引力测度，如用地规模等；

C_{ij}^k 为 i 与 j 区域之间的离散性测定，一般可定义为

$$C_{ij}^k = a_1 t_{ij}^k + a_2 e_{ij}^k + a_3 d_{ij}^k + P_j^k + \delta^k$$

式中：t_{ij}^k 为 i 到 j 用 k 种交通工具的出行时间；

e_{ij}^k 为 i 到 j 用 k 种交通工具出行的附加出行时间，如等车、停车；

d_{ij}^k 为 i 到 j 用 k 种交通工具的出行距离；

P_j^k 为 j 区域的附加花费如停车费等；

δ^k 为其他花费；

a_1、a_2、a_3 为比例因子。

$f(C_{ij}^k)$是关于 C_{ij}^k的函数，可以有如下形式：

$$f(C_{ij}^k) = (C_{ij}^k)^{-2}$$

或

$$f(C_{ij}^k) = e^{-\beta C_{ij}^k}$$

O_i，D_j 的数值可以是 i 地区和 j 地区所选择的特征数值的函数。

如：D_j 为交通发生模型中的 $T_i = \sum_i r_{ij}$

$$D_j = \sum_{m=1}^{m} b_m r_j^m$$

r_j^m 可以取就业人数，占地规模等。

计算出 $L \times m \times n$ 个 T_{ij}^k的值，并在图上表示出，即可反映出城市交通流的流量、流向分布情况。

5.3　交通量最优分配规划

交通量在交通网络中的最优分配，对于客流，往往采用最短路径算法，以出行距离最小为原则，求出各居住小区到各出行目的地的出行量。对于货流，一般采用线性规划中的运输模型。

1. 平衡运输模型与不平衡运输模型

以 X_{ij}表示待求的从地点 i 到地点 j 的最佳运输量并由模型求解得出。D_j 表示 j 地点所需要的货物到达量，S_i 表示 i 地点货物可供数量或生产量，C_{ij}表示从 i 地点运输到 j 地点单位产品运输成本。

给出以下约束：

生产量约束：$\sum_j X_{ij} = S_i$

消费量约束：$\sum_i X_{ij} = D_j$，$X_{ij} \geqslant 0$

目标函数，总运费最省：

$$Z = \sum\sum C_{ij} \cdot X_{ij} \rightarrow \min$$

由于假定生产量严格地等于消费量，所以称为平衡运输模型。

如果：$\sum_j X_{ij} \leqslant S_i$

$$\sum_i X_{ij} \geqslant D_j$$

$$Z = \sum_i \cdot \sum_j C_{ij} X_{ij}，X_{ij} \geqslant 0$$

即要求全部消费量小于生产量；全部生产量大于消费量，则称为不平衡运输

模型。

这是一个线性规划模型,采用单纯法即可求解之。

2. 交通量分配模型

是以引力模型为基础的分配模型,在建立该模型时需要考虑下述因素:

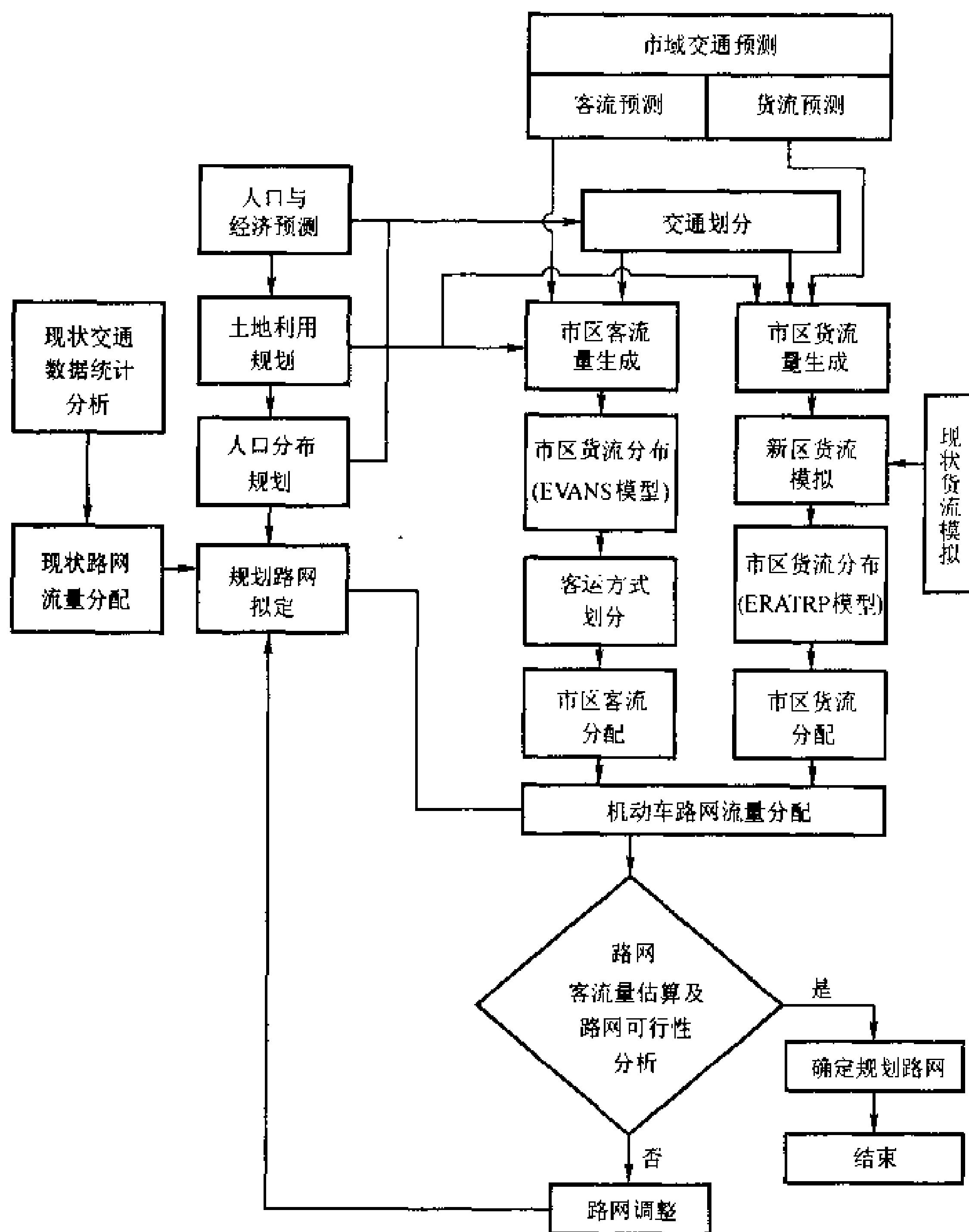


图6-16　城市综合交通规划定量分析的流程框图

(1) 道路功能的合理划分。在城市中,道路按功能可以分为过境道路、主要

交通干道、次干道、主要生活性道路、次要生活性道路和联系城区主干道和过境公路的进出城干道。

主要交通干道是流量分配的基本计算对象，主要生活性道路是联系居住与就业的主要道路，也是流量分配的重要对象。

(2) 交通流对速度的要求。除了 O ~ D 流量在空间上的分配之外，还应保证速度要求。为了达到速度要求，可限定交通流方向，如采取单行道；或限制流量，如制定限制高峰流量或在一定时间限制某一类的车流通过等。

(3)道路使用的经济性。在优化过程中，如果仅按运费原则，可能使部分道路流量过稀而引起不经济，同时也会使部分便捷通道超负荷，为此，应用约束条件表示充分利用道路及其设施的要求，保证城市道路各尽其能。

(4)对交叉路口流量的限制。由于主干道各交叉口往往流量十分集中而造成堵塞，因此，有必要在模型中对交叉口各方向的流量作出约束，以保持交叉口的畅通。

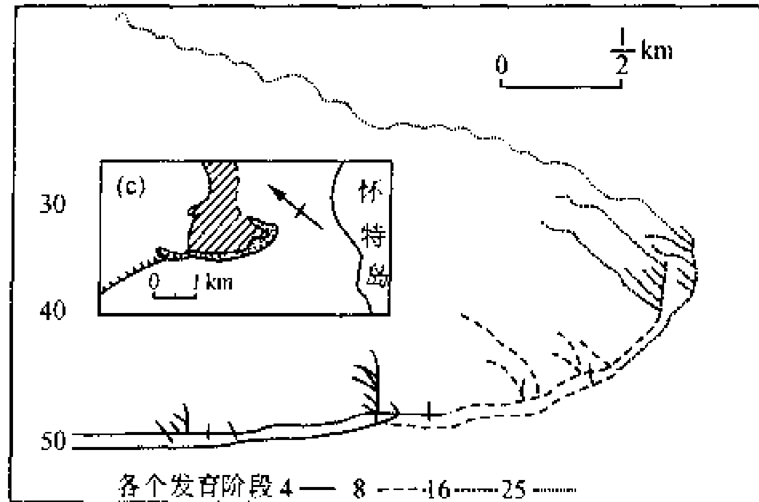
以上几个方面都是建立模型时应考虑的约束条件。

在空间型信息系统支持下开展城市综合交通规划分析，是以交通工程理论为依据，进行现状交通调查数据的统计分析和远期交通流的模拟预测，为确定城市用地布局和优选规划路网提供定量的决策依据。其基本工作流程如图 6 - 16 所示。

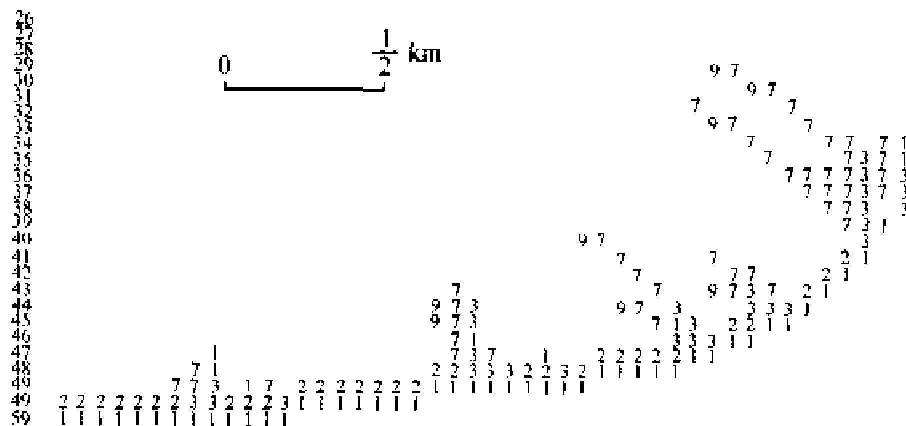
§6 地学模拟模型

地学模拟模型是应用计算机、数字模拟技术及综合分析的方法来模拟许多地理过程或现象，例如气候变迁、沙漠化过程、土地退化过程、湖泊沼泽化、河道冲淤、沙嘴发育(图 6 - 17)，以及土壤侵蚀等，使得受几个因素的共同影响，要经过若干年才能完成的地理过程，采用计算机模拟模型，只需几分钟就能得出类似的结果，为资源开发、国土整治、水土保持、工程论证等提供依据。以下，以土壤侵蚀的模拟模型为例，介绍应用地理信息系统技术研究土壤侵蚀量的方法。

利用地理信息系统的数值分析方法来估算土壤侵蚀量，是首先确定土壤侵蚀的数值分析模型，根据模型确定影响土壤侵蚀的因子，这些因子必须能够反映不同的土壤性质，不同的坡面形态，以及不同的植被条件等，然后选择格网尺寸，建立各个影响因子的栅格数据文件，最后将多种信息加以复合，确定研究地区土壤侵蚀量的各种不同等级，为制定区域的水土保持规划提供依据。



(a) 英格兰南海岸卡斯特尔沙嘴的实际形态



(b) 计算机模拟的沙嘴模型

图 6－17 沙嘴发育的模拟模型

（据 King 等，1978）

6.1 确定土壤侵蚀的数值分析模型

土壤侵蚀的数值分析模型随具体区域而不同，美国普渡大学曾根据 30 余个观测站的数以万计的资料，用电子计算机加以分析，得出下列通用的土壤流失方程。

$$A = 0.224 \cdot R \cdot K \cdot L \cdot S \cdot C \cdot P$$

式中：A 为土壤流失量，$kg \cdot (m^2 \cdot a)^{-1}$；

R 为降雨侵蚀力；

K 为土壤可蚀性；

L 为坡长；

S 为坡度；

C 为植被覆盖度；

P 为土壤侵蚀控制措施。

6.2 设计土壤侵蚀数据处理流程

根据模型确定的土壤侵蚀因子，研究各个因子的计算或提取所根据的数据源和方法、数据组织和编码方式，然后拟定具体的数据处理流程(图 6 - 18)。

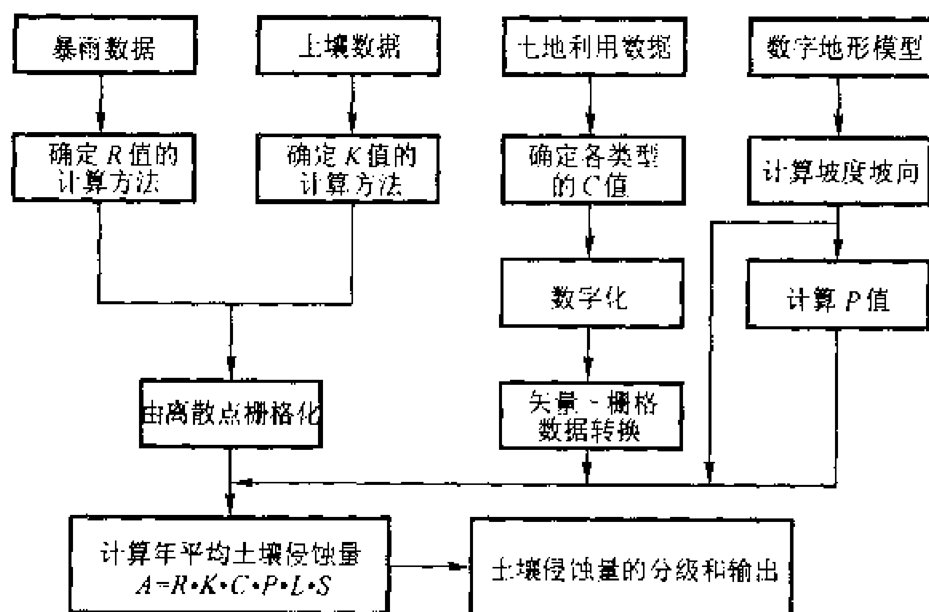


图 6 - 18 土壤侵蚀数据处理流程图

6.3 土壤侵蚀图的输出

根据计算土壤侵蚀贡献的公式，将各网格的土壤侵蚀量换算为土壤侵蚀贡献量。流域内各网格土壤侵蚀贡献量之和等于流域年均产沙量，并应等于流域出口断面实测的年均输沙量。在求取流域年均产沙量前，首先要提取流域边界，然后将流域边界文件与土壤侵蚀量的栅格数据文件进行叠合，结果流域边界外的栅格值均为零，流域内的栅格值被保留，这样计算流域产沙量时不再受流域外数值的影响，最后将栅格的土壤侵蚀贡献量，按照拟定的分级方法，并且不同等级的贡献量以不同色调的符号表示(图 6 - 19)。如果根据试验区的土地利用方式、土层厚度、土壤性质和降水特点，确定区域的土壤流失容许量，例如设 $T = 0.8\ \mathrm{kg \cdot (m^2 \cdot a)^{-1}}$，则根据区域的年平均土壤侵蚀量减去土壤侵蚀容许量，结果大于零的栅格，表示其土壤侵蚀已超过容许限度，便得到土壤侵蚀超限区域分布图。这两种地图对于确定流域的主要产沙区，明确流域水土流失治理的重点区

域,以指导区域的水土保持规划,具有重要的指导意义。

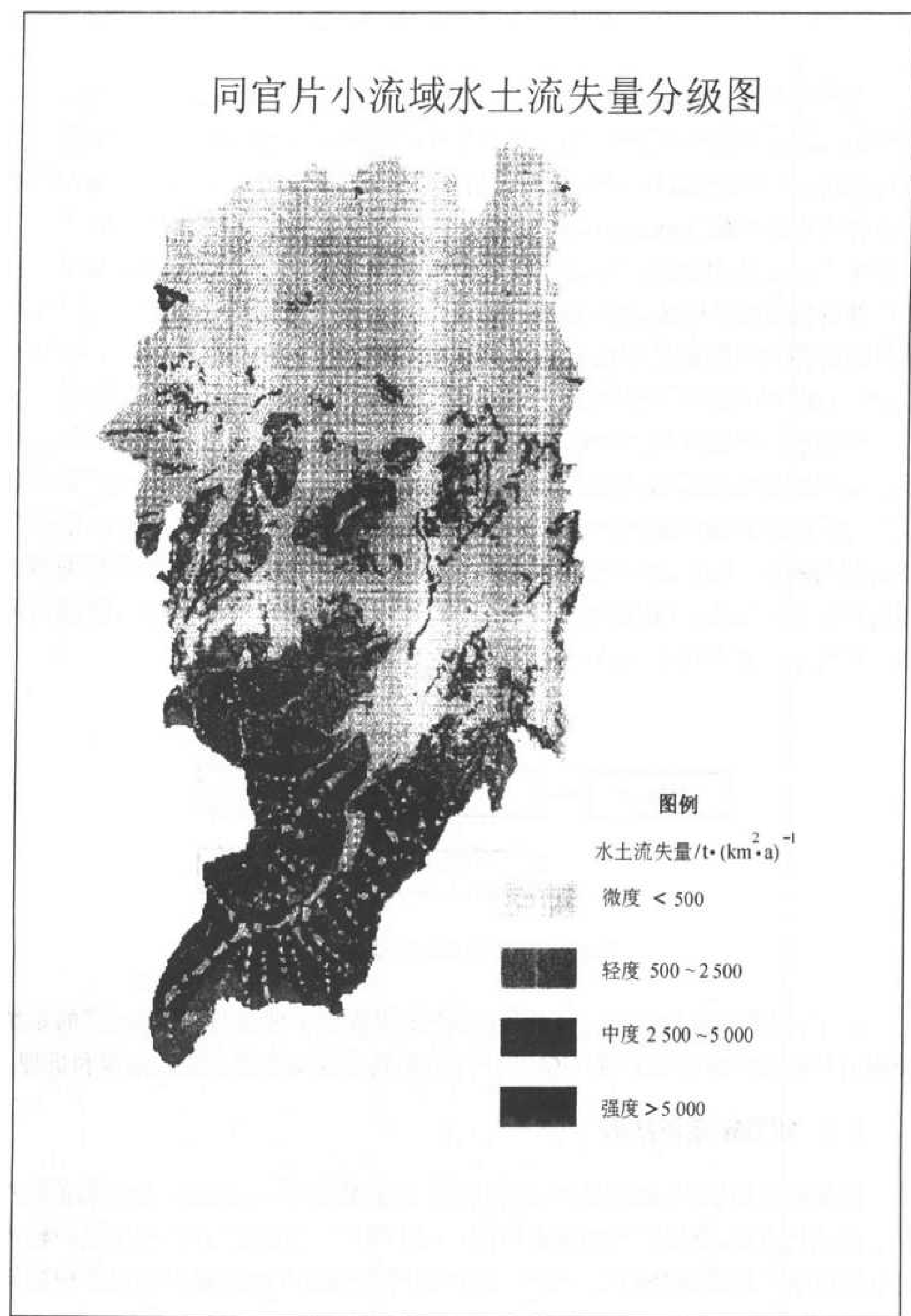


图 6-19 流域水土流失分级图

§7 专家系统概述

专家系统是研究模拟人类专家的推理思维过程，将领域专家的知识和经验，以知识库的形式存入计算机，系统可以根据这些知识，对输入的原始事实进行复杂的推理，并作出判断和决策，从而起到专门领域专家的作用。具有这种功能的系统称为专家系统（expert system）。将专家系统技术应用于地理信息系统领域具有重要的意义，因为地理信息系统常常要解决多层次、多因素、时变型和具有非线性变化的地学问题，解决这类问题是一项具有一定创造性的过程，这里，专家的经验和知识起着重要的作用。据估计，目前全世界已开发的专家系统有2 000～3 000种，它们广泛应用于疾病诊断、地质勘探、地学分析和军事领域。

专家系统的基本构成如图6－20所示。首先，对某一领域具有深入研究和专门知识的专家，将其知识变为可供计算机使用的形式，这个过程称为“知识获取”。然后，将获取的知识转换为一系列辨认或描述实体的规则，并存入系统存取的知识库中。最后，用户通过界面进行询问，推理机构便利用知识库中的知识和用户输入的信息，完成推理和生成结论，并可根据用户的要求给出相应的解释。这就是专家系统工作的一般原理。

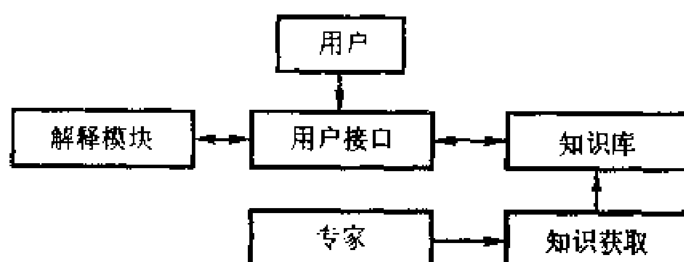


图6－20 专家系统的基本结构

以下，以基于南京大学肖楠森教授的“新构造控水理论与找水方法”的系统知识而研制的找水专家系统（NCGW）[①] 为例，具体说明专家系统的结构和机理。

7.1 NCGW系统结构

该系统由知识库、推理机构、解释系统、动态数据库、人机接口和知识获取六个主要模块组成，其模块间的关系如图6－21所示。系统运行时，根据用户输入的该地区水文地质条件和可信度值，推理机构利用输入的数据和知识库中的知

① NCGW是Neotectonic Control Ground－Water Expert system（新构造控水专家系统）之缩写。该系统由南京大学计算机科学系研制。

识,推断出该地区是否有可开采的地下水。例如,如果该地区为丘陵岗地,同时又为新构造断裂带,系统便进行新构造断裂带富水性的测试,同时推导出井深、井位、打井方法等结论。

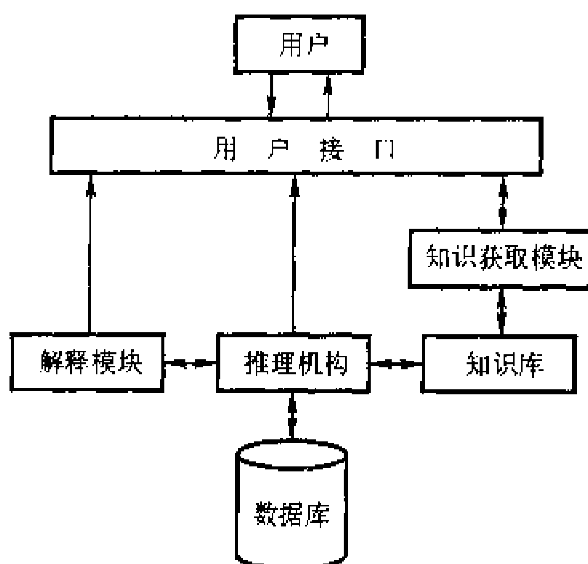


图 6-21 NCGW 专家系统的基本结构

7.2 知识库格式

知识库存储着特定领域内的大量事实和规律,是专家实践经验全面和真实的体现。知识库中知识的表示与组织是系统的基础,关系到系统的推理是否能够模拟出专家的思维过程。在 NCGW 系统中,采用产生式规则的知识表达方式(图 6-22)。每个产生式包含一个"情况-行为"对,在推理系统中,情况就是前提,行为就是结论,因此知识的单位是:

```
nfdv(['de','da','1'],5)
nfdv(['de','da','2'],5)
nfdv(['de','da','0'],0)
nfdv(['de','nc','1'],5)
nfdv(['de','nc','2'],5)
nfdv(['de','nc','0'],0)
nfdv(['de','nb','1'],5)
nfdv(['de','nb','2'],10)
nfdv(['de','nb','3'],5)
nfdv(['de','nb','0'],0)
nfdv(['de','nbi','1'],10)
```

续表

```
nfdv(['de','nbi','2'],10)
nfdv(['de','nbi','0'],0)
nfdv(['de','nbg','1'],10)
nfdv(['de','nbg','2'],10)
            ⋮
            ⋮
           (1)
rule(['de','da','1'],'该走向的为断裂倾向?')
rule(['de','da','2'],'该走向的为断裂倾向是南南西方向')
rule(['de','da','0'],'该走向的为断裂倾向是北北东方向')
rule(['de','nc','1'],'该走向的为断裂倾向中有什么沉积物?')
rule(['de','nc','2'],'断裂中有纤维状方解石')
rule(['de','nc','0'],'断裂中有梳状方解石')
rule(['de','nb','1'],'该走向的断裂中有什么充填物质?')
rule(['de','nb','2'],'断裂中有黄褐色黏土')
rule(['de','nb','3'],'断裂中有灰褐色黏土')
rule(['de','nb','0'],'断裂中有白色黏土成白色粉状')
rule(['de','nbi','1'],'该走向的断裂面上有何种物质分布?')
rule(['de','nbi','2'],'断裂面上有褐色铁锈')
rule(['de','nbi','0'],'断裂面上有褐色铁锰物质细脉')
rule(['de','nbg','1'],'该走向裂隙中有什么风化褐色现象?')
rule(['de','nbg','2'],'裂隙风化现象是灰黄色')
```

图 6-22 NCGW 系统知识库示例

If〈前提条件〉Then〈结论〉

NCGW 系统的知识库中存有 400 条产生式规则，根据实际应用的需要，可随时添加或修改知识库中的规则。知识库中规则的内部表示格式为

谓词名([父节点名,子节点名,菜单选择项编号],可信度值或规则体)

各规则的结论部分附加一定的确信度，推理后得出的结论，如果确信度满足一定的值，即认为结论成立，否则结论就不可信或不成立。

7.3 推理机制

知识表示是专家系统赖以生存的基础，而推理控制策略则是系统的灵魂，它类似于人类专家运用来源于知识和实践的一般思维规则。NCGW 系统采用反向推理策略，即系统从代表目标的谓词公式出发，通过搜索一组规则证明事实谓词

公式成立，在规则调用过程中，动态地生成上下文的与/或树。图 6-23 表示部分的与/或树。

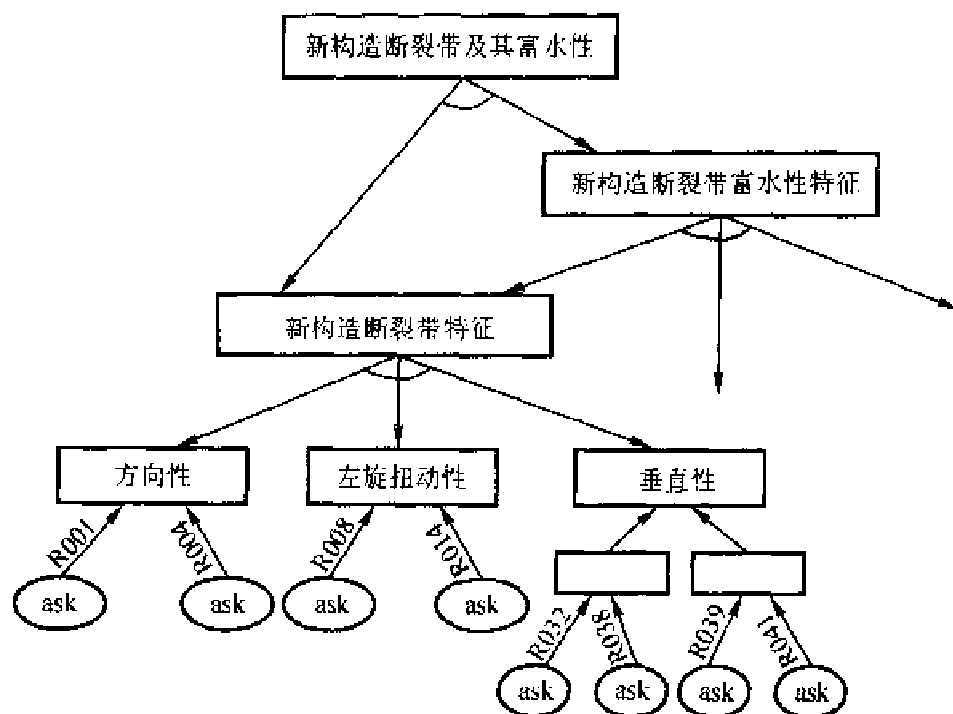


图 6-23 NCGW 系统反向推理树示例

7.4 知识获取

NCGW 系统的知识获取通过机器与专家的对话来完成，系统以菜单方式(图 6-24)主动向专家提问，以获取知识，包括建立、添加、删除或修改规则。例如，如果需要添加一条规则，可用中西文直接键入这条规则的内容，系统自动将这条规则转化为内部形式存入知识库，同时将对应规则名排入推理控制流中，并自动匹配生成对应菜单。添加规则的控制流程图如图 6-25 所示。

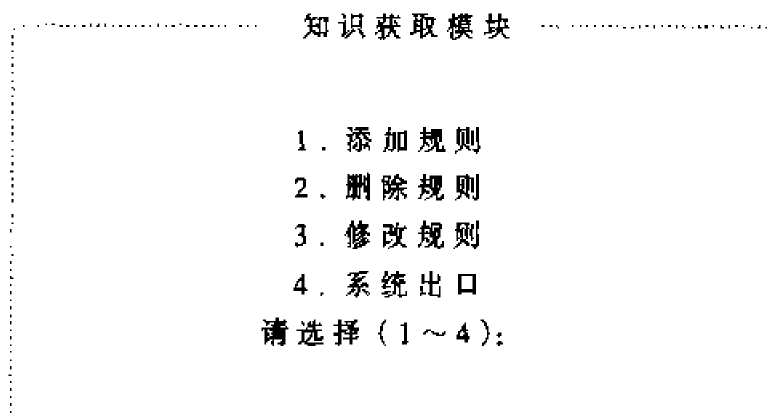


图 6-24 知识获取菜单画面

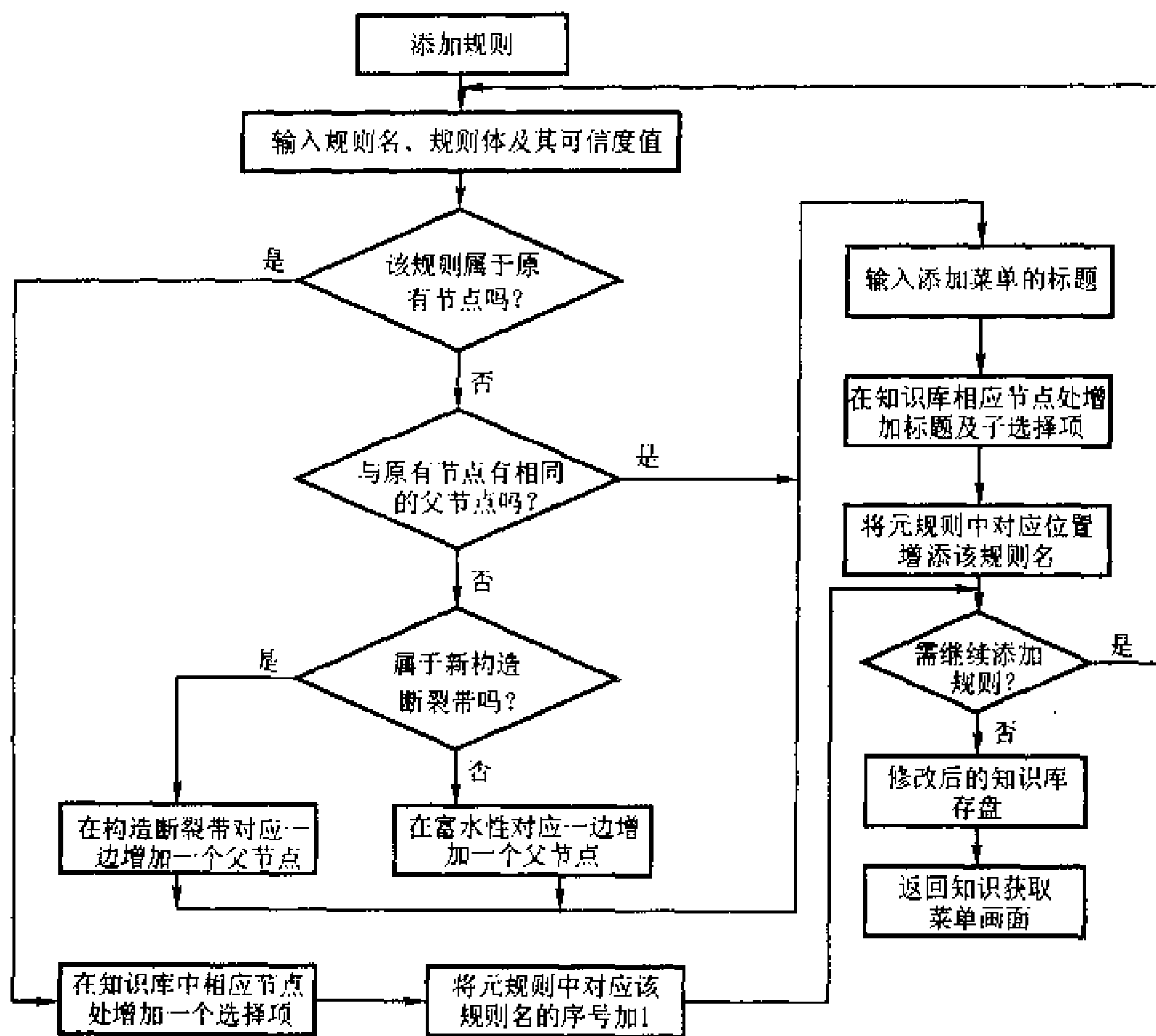


图 6-25 添加规则的控制流程图

思考题

1. 结合你的专业领域(水土流失、洪水灾害、土地分等定级等),叙述基于 GIS 的地学应用模型的建模步骤和方法。

2. 通过应用实例或实际操作过程图,说明在解决地学问题中,如何将专业知识与 GIS 的应用过程进行有机的结合,并讨论两者之间的相互关系。

3. 为公园选址,要求该公园的位置必须是从主要公路上容易到达,要减少噪声干扰,同时依山傍水,而且公园的可利用面积为最大。试以 GIS 方法模拟选址分析步骤,设计其数据处理流程图,并说明其有关的操作。

4. GIS 支持下的建设项目环境管理与分析(选自 2000 年国产 GIS 基础软件测评试题)。

背景条件:现有一经济开发区,它由五大功能区组成,各功能区范围地理坐标和允许排污负荷量如表 1、表 2 所示。现在此经济开发区内有两个建设项目(其地理坐标如表 3 所示),它们的设计排污量负荷分别为 5 t/km^2 和 8 t/km^2。

环境管理审批条件：建设项目通过环境审批的条件是建设项目设计排污量负荷不超过建设项目所占功能区允许排污量负荷之和。

题目要求：

(1) 编制专题地图：

① 开发区功能区图，显示出各功能区的面积、允许排污负荷量(要求有图例)；

② 建设项目图，显示建设项目及其与开发区各功能区的空间位置关系。

(2) 输出统计表：

① 建设项目使用开发区各功能区土地面积统计表；

② 建设项目在开发区各功能区的排污负荷量统计表；

③ 项目审批结果表(表格形式可按环境项目审批条件自行设计)。

(3) 建设项目可行性分析：

对于不可行的建设项目，可通过移动其地理坐标，使之满足环境管理审批条件。试给出在地理坐标移动最小的前提下的解决方案。

(4) 构建可完成上述要求(1)、(2)、(3)任务的应用界面，使之能够适用于其他建设项目环境管理。

表 6-9 功能区坐标

A 区		*B* 区		*C* 区		*D* 区		*E* 区	
X	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*	*X*	*Y*
0	0	0	6	0	2	2	0	8	0
2	0	4	10	2	2	8	0	10	0
2	2	0	10	6	4	6	4	10	10
0	2	0	6	4	10	2	2	4	10
0	0			0	6	2	0	8	0
				0	2				

表 6-10 各功能区允许排污负荷量

功能区	允许负荷量/t·Um^{-2}
A	2.0
B	8.0
C	6.0
D	5.0
E	10.0

表6-11 两个新建项目地理坐标

项目1		项目2	
X	Y	X	Y
1	1	4	6
3	1	8	6
3	4	8	2
1	4	4	2
1	1	4	6

第7章 地理信息系统的设计与评价

§1 GIS设计概述

1.1 系统设计的目的

地理信息系统,按其功能和内容,可以分为工具型地理信息系统和应用型地理信息系统。这里的系统设计是指应用型地理信息系统的设计。所谓应用型地理信息系统,是指在工具型或基础型地理信息系统的基础上,经过二次开发,建成满足专门用户解决一类或多类实际问题的地理信息系统。因此,应用型地理信息系统的主要特点是,它具有特定的用户和应用目的,具有为满足用户专门需求而开发的地理空间实体数据库和应用模型,它继承工具型地理信息系统开发平台提供的大部分功能和软件,以及具有专门开发的用户应用界面等。

应用型地理信息系统,根据其应用层次的高低,又可分为空间事务处理系统(STPS)、空间管理信息系统(SMIS)和空间决策支持系统(SDSS)。STPS的主要目的是通过应用GIS的数据库技术,实现由传统的事务处理向计算机处理的转换,例如在房产、地籍等部门有着广泛的应用。SMIS的主要目的是实现空间信息管理的高效率、模型开发和空间数据的动态更新,其功能不但表现为数据的查询和统计,还具有专业模型的分析应用等功能,它在城市规划、土地利用、道路交通管理、管网规划管理等领域具有广泛的应用。SDSS主要用以解决半结构化和非结构化的决策问题,除了需要利用地理信息系统的数据库和空间分析技术,模型库及其管理系统是决策支持系统的核心,它在宏观决策、行业发展规划等领域具有广泛的应用需求。

建立这些应用型的地理信息系统,要求功能能满足需求,系统运行稳定可靠,系统应用能达到高效益,实现业务操作的手工模式向信息化模式的根本转变,以便提高管理和决策的高效率和科学化。

1.2 系统设计的模式

地理信息系统最早的设计模式,是Calkins在1972年,由国际地理学会地理

数据收集和处理委员会主持召开的地理数据处理学术会议上提出来的，后来又经过了几次修改和补充。这个最早的设计模式称为结构化的系统设计模式，由四个组成部分构成(图7-1)：①通过访问用户，调查用户的需求和数据源，确定系统的目的、要求和规定；②描述和评价与系统设计过程有关的资源和限定因素，例如现有的硬件、软件和有关的政治、法律因素等；③说明和评价所拟定的不同系统，这些系统能够满足所规定的要求；④对拟定的系统作最后的评价，从中选择一个运行的系统。

该模式的主要特点是强调对用户的调查和系统功能需求的分析。在系统设计的各个阶段都要写成有关的文件，以便进行评价，以及用户要参与系统的设计，以免系统设计的失误。

自从这个最早的地理信息系统设计模式诞生以来，地理信息系统的开发已经取得很大的进展。原来的设计模式是假定系统的大部分组成(除硬件外，包括所有的软件和数据库)都需要由系统设计人员来完成，有时甚至包括处理空间数据的某个专门的硬件。现在的情况不同了，不但有许多处理空间数据的重要软件，而且有现成的系统和空间数据库，因此需要对原来地理信息系统设计模式进行修改，修改后的地理信息系统设计模式如图7-2所示。其主要的设计思想，是强调对现有的各个组成部分，包括硬件、软件和数据库，进行深入、认真的评价，以研究其满足系统功能的程度，保证所设计系统的实用可靠，及有效地处理数据和使用周期长等要求。该模式采用了管理信息系统和软件工程的一些设计理论，包括：

(1) 目的与任务。每个系统都要对目的和任务作详细的说明，指出该系统的目的，谁是主要的用户，以及如何使用该系统。关于任务，要说明所要完成的工作，以及总体评价所采用的方法，目的和任务的说明要非常详细，以便用户进行评论和评价，而且这种说明代表着用户和系统设计人员对话的开始，并且在系统设计的过程中还要继续进行这种对话。

(2) 概念的定义。介绍系统的各个主要组成部分，分别按照输入、输出、主要过程和数据库来说明系统的基本结构，包括主要模块、系统开发的主要资源、主要的限制条件等。

(3) 功能的要求。具体说明该系统要做什么，对每一种功能要求都要说明，包括功能的技术特征、功能的目的、具体的标准和满足的条件等等。功能的要求一般分为以下三类：①该系统必须具有的功能；②如果时间和资源条件允许，希望具有的功能；③其他有意义的功能。

实际上只有第一类功能才是系统的真正要求。每一种功能要具体规定：①输入(来源、数据、频率)；②输出(格式、数据量、用户)；③功能需要的处理步骤；④功能成功地实现所需要的条件；⑤功能生成的数据。

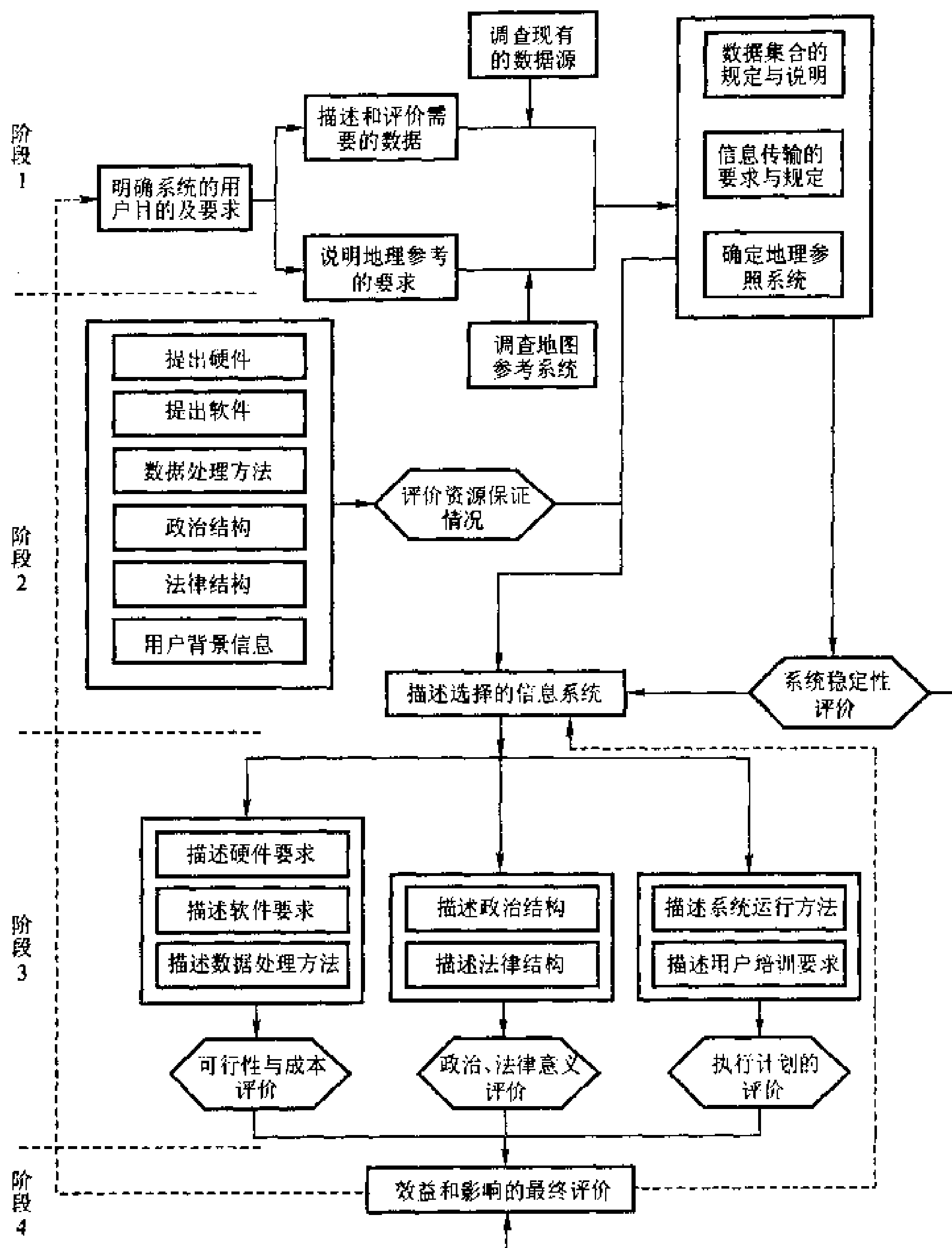


图7－1 结构化的系统设计模式

（据 H.W.Calkins 修改）

除了规定要完成的功能外，还要说明该系统期望的性能和特征，质量控制措施，以及该系统与其他部分的接口等。

显然，在系统开发的过程中，要求可能发生变化，因此要制定专门的计划进

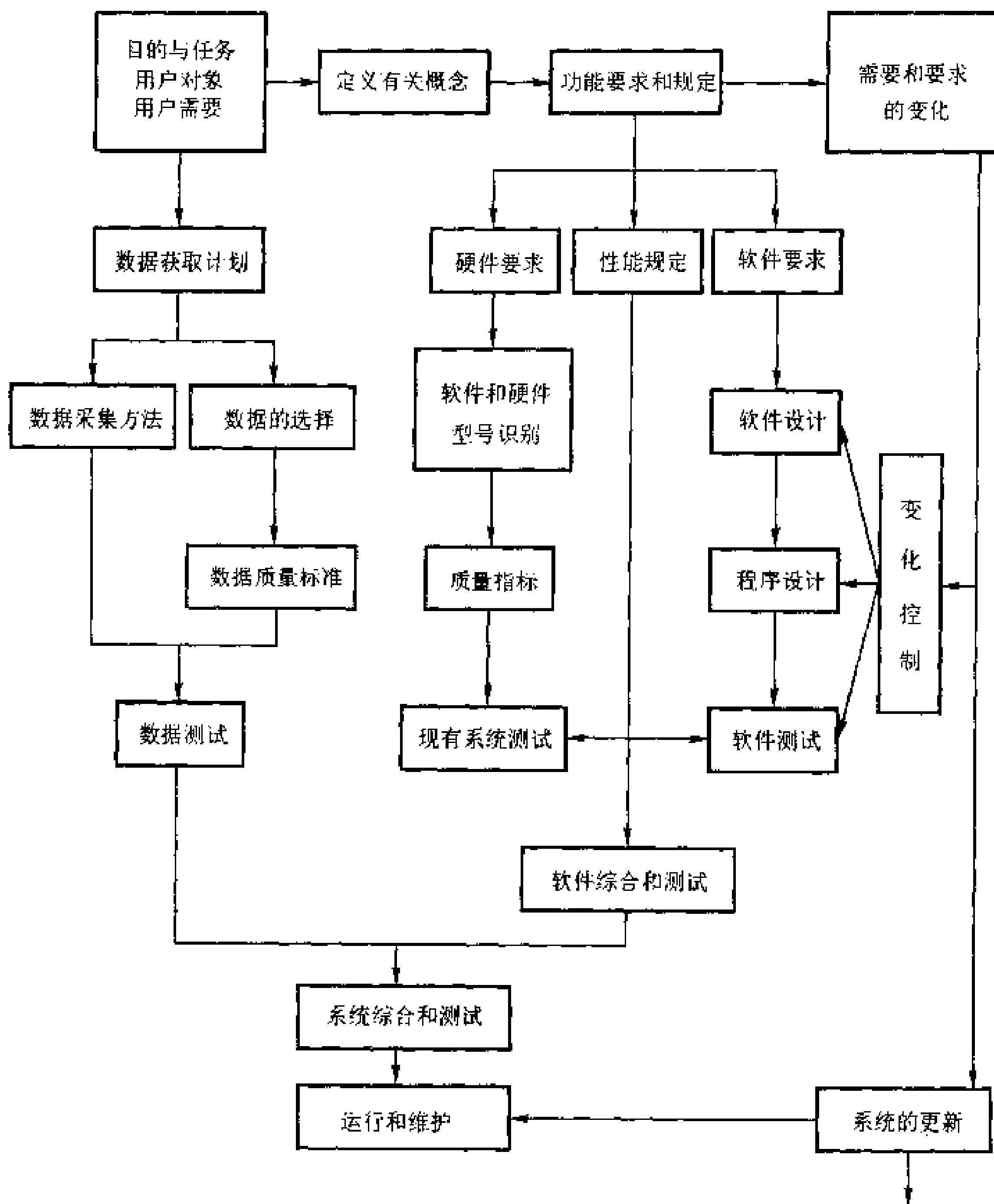


图7-2 结构化的系统评价模式
(据 H.W.Calkins 修改)

行处理。

(4) 性能测定。在系统设计过程中,要对各个组成部分分别进行测试,对综合以后的整个系统要进行最后的测试。具体测试的内容包括:硬件、软件模块、数据库的质量控制等。测试根据所说明的功能要求和规定的标准进行,测试应考虑以下各种条件:①系统的正常操作条件;②重点测试,包括最坏情况和极端

操作条件;③逻辑测试,指检查各种可能的逻辑条件;④线路测试。

硬件、软件和数据库的测试,是对系统进行总体评价的最后阶段,那么,在系统设计的各个阶段,谁进行这种评价。显然,在目的、任务和要求的评价中,与用户有关。而对系统设计其他方面的评价,则需要其他有关的技能,一个系统的有效性取决于软件的质量,而对软件的评价,则必须由具有软件工程专门知识的专家来进行。

以上是 H.W.Calkins 于 1984 年提出的地理信息系统开发的基本模式,这种模式是软件工程中的结构化思想在应用型 GIS 开发中的应用。它针对系统开发中存在的用户需求不明、开发过程无序、开发人员之间缺少有效交流等几个主要问题,通过一系列规范性的支持技术和方法,力图将系统开发这种只存在于开发人员头脑中的抽象思维活动,转化为一组明确的、可控制的和可以检查的具体任务,要求做到对用户需求的充分了解和定义,严格有序的开发过程和组织管理,强调开发人员与用户之间的交流和协作,重视系统建设文档的完整和规范,以及加强系统开发综合的质量保证措施和计划等,最终实现以较少的人员、时间和投资,开发出尽可能符合用户要求的、高质量的应用系统的目的。

1.3 系统设计的流程

应用型地理信息系统的设计,大致可以分为四个主要阶段,即系统分析、系统设计、系统实施、系统运行与维护(图 7-3)。这四个阶段构成了系统的生命周期,如果系统设计按阶段进行,预先规定每一阶段的开发目标和任务,然后按照一定的准则顺序开发实施,这种方法被称为生命周期设计法。由于生命周期法严格分阶段进行,便于开发工作的组织管理,但是也存在许多缺陷,例如系统开发过程比较长,新系统实际效果的可见性差,来自最终用户的反馈比较迟,不容易把握用户需求的变化等。为此人们设法寻求另外一种新的开发思路,称为原型化设计法。

所谓原型化设计法,就是根据用户提出的需求,由用户与开发人员共同商定其中重要和基本的开发目标,然后选择一个试验区,设计出初步方案,在较短时间内开发出一个能满足用户基本需求的初步原型或系统雏形,交用户试用,经过一段时间的运行后,根据用户意见对原型加以修改或扩充,产生一个新的原型版本,如此反复和迭代,最后形成一个比较完善和质量较高的应用型地理信息系统。原型化设计法的开发过程如图 7-4 所示。原型法与生命周期法相比较,原型法能使用户更积极地参与新系统的设计和实现,能使开发人员及时获得用户的反馈意见,能更快地看到系统的实际效果,能使系统开发的风险降到最低限度。但一般认为,当用户对于新系统的功能需求十分明确时,系统设计可直接采用生命周期法,而不应该采用原型法,以利于开发经费和开发时间的控制。

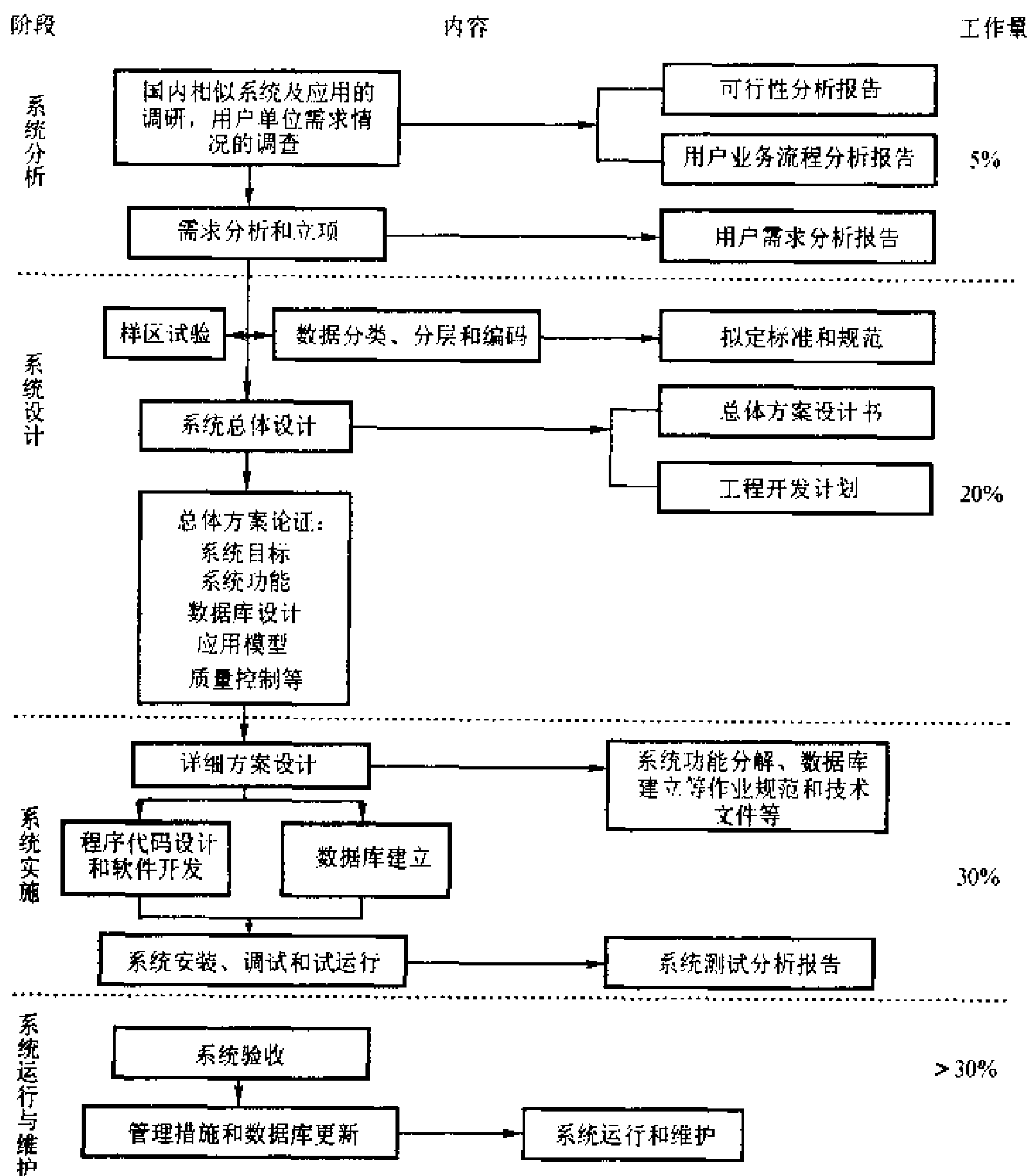


图7-3　地理信息系统设计流程图

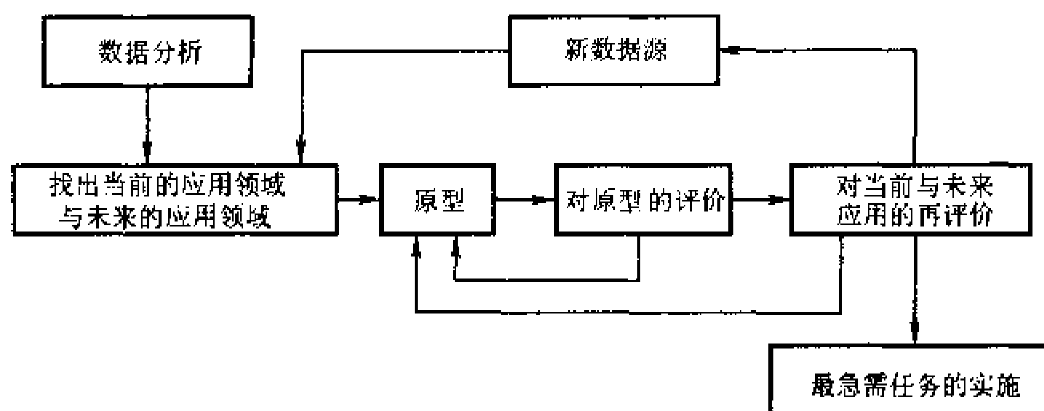


图 7-4 原型化设计法开发过程图
(据熊允泰,1995)

§2 地理信息系统的设计

2.1 系统分析

系统分析(system analysis)一词源自美国的兰德公司(RAND: Research And Development),其基本思想是从系统观点出发,通过对事物进行分析与综合,找出各种可行的方案,为系统设计提供依据。它的任务是对系统用户进行需求调查和可行性分析,最后提出新系统的目标和结构方案。系统分析是使设计达到合理、优化的重要步骤,其工作深入与否,直接影响到将来新系统的设计质量和实用性,因此必须予以高度重视。

用户需求调查即调查系统用户对开发的 GIS 系统的功能要求和信息需求情况。具体调查的主要内容有:

·Who,谁使用该系统,该系统的用户结构如何,哪些是直接用户,哪些是间接用户,哪些是最终用户,哪些是潜在用户。当前本部门的组织机构、人员分工和职能情况,现有的业务流程和工作效率等。

·What,新系统是做什么用的,它需要具备哪些功能,它应能解决和处理哪些类型的问题,因此需要具有哪些设备、资源、数据等。

·Why,为什么需要具有这些功能和条件,具有这些功能以后与常规的业务流程有哪些不同点和优越性,对现行系统和建立的新系统从功能、效率、效益等方面做详细调查及对比研究等。

·Where,建立新系统所需要的资源从哪里获取,特别是数据资源能否得到保障,以及解决系统硬件和软件的途径等。

·Quality,指具体的技术指标、性能要求和可靠性要求,例如数据精度、运行速度、系统安全保障机制等,要认真听取用户的意见和要求。

用户调查一般采用访问、座谈等方法。在调查前,应拟定出需求调查提纲。在调查中,重点应弄清用户对所要开发系统的功能、数据内容、应用范围等方面的要求,并详细考察用户原来的业务范围、工作流程以及部门之间的联系等。在调查后,需撰写用户需求调查报告,内容包括:用户对系统的要求,用户目前的业务范围、工作流程和存在的问题,可用的数据源情况,现有的技术力量、设备条件等。同时,要对这些需求进行可行性分析,着重从社会、技术和经济三大要素分析开发新系统的可行性,确定哪些需求可以实现,哪些需求需要调整和简化,哪些需求作为近期目标或远期目标等。用户需求和可行性分析报告,这是系统设计的重要依据,要用文字和图表详细阐述。用户需求和可行性分析报告经过审批,表示系统开发项目得到立项,才能转入系统设计阶段。

2.2 系统设计

1. 总体设计

总体设计又称为逻辑设计,其任务是根据系统研制的目标来规划系统的规模和确定系统的各个组成部分,说明它们在整个系统中的作用与相互关系,以及确定系统的软硬件配置,规定系统采用的技术规范,并作出经费预算和时间安排,以保证系统总体目标的实现。最后撰写系统总体设计方案,作为重要的技术文件提供论证和审批。

总体设计的主要内容包括:

(1) 用户需求。阐明系统的用户构成、不同用户对系统的要求、系统应具备的功能等。

(2) 系统目标。阐明该系统的应用目标,属于演示系统或运行系统、单机运行系统或分布式运行系统、事务处理系统或信息管理系统等。

(3) 总体结构。根据系统功能的聚散程度和耦合程度,将系统划分为若干子系统或功能模块,构成系统总体结构图。例如深圳市规划国土管理信息系统的总体结构(图7-5)。

(4) 系统配置。指系统运行的设备环境,包括计算机、存储设备、输入和输出设备以及网络等,并说明其型号、数量和内存等性能指标,画出硬件设备配置图(图7-6)。软件包括计算机系统软件、网络管理软件、地理信息系统基础软件、数据库管理系统软件、应用软件等,并说明其版本、数量和性能特点。

系统配置应遵循技术上稳定可靠、投资少、见效快、立足现在和顾及发展的原则。技术上稳定可靠是指采用国内外经过实践检验证明其为成熟的硬件和软件,同时以满足本系统的技术和性能指标为准则,不单纯追求最高档设备与昂贵

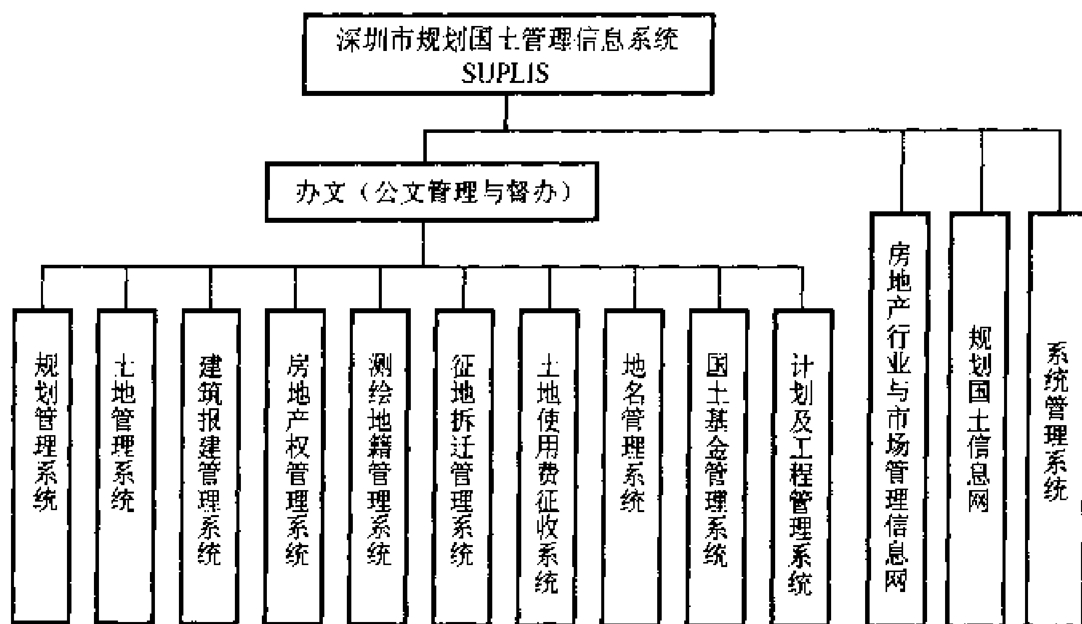


图7-5 深圳市规划国土管理信息系统总体结构

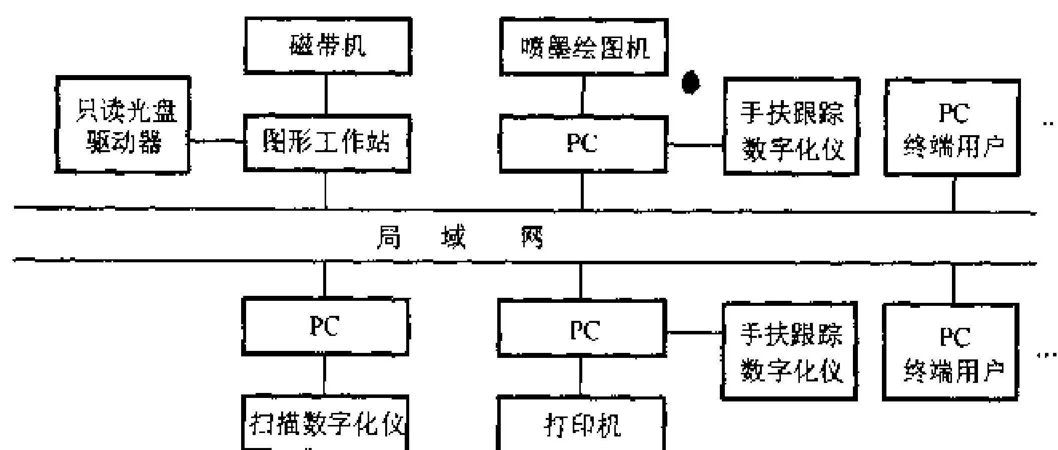


图7-6 系统硬件配置图

的软件；投资少、见效快即根据经济实力和技术力量，选择合适的配置，能较快地收到实际效果；立足现在、顾及发展是指应以完成目前的要求为主，并顾及系统的可扩充性和将来的发展。由于系统具体目标和服务范围不同，系统配置方案也有很大差异。例如多用户的系统，为实现资源共享、协同工作和并行处理，这时客户/服务器结构、分布式数据库和网络化等配置方案，便是该类系统的基本要求。因此，系统配置方案的确定具有动态变化的特征。

(5) 数据库设计。数据库是系统的核心组成部分，一个系统可以具备一个或多个数据库。按信息内容，可将数据库分为基础数据库和专题数据库。按数据类型，可将数据库分为空间数据库和属性数据库。数据库设计要确定空间数

据与属性数据的管理模式，集中式或分布式的建库方案，采用的数据结构类型和数据库管理系统以及数据分类等。

(6) 系统功能。由于应用型地理信息系统继承了开发平台所提供的大部分功能，因此应用型地理信息系统功能设计的任务不在于解决基本功能，而在于解决用户所需要的特定功能。例如，以一个地下管线信息系统为例，其系统功能包括：①数据录入与查错，实时地将普查数据转换为满足地下管线信息系统要求的数据，以保障地下管线的综合管理与应用。②综合查询与统计，提供按图号、道路名、单位名查询任意范围的管线，并统计计算各类管线的长度、面积、体积等。③网络分析和诊断，分析各类地下管线发生事故或故障(如漏水、漏气等)时，影响区域的范围，涉及到哪些阀门需要关闭和维修等。④断面生成与分析，断面分析是道路与管线工程规划设计、管理的基础，也是地下管线工程综合的主要依据。断面分析分为纵断面与横断面两种，系统应能生成和分析任意位置和方向的横断面以及生成和分析连续管线的纵断面。⑤管线工程辅助设计，以国家有关管线工程的最小覆土深度、管线最小水平净距、管线交叉时的最小垂直净距等规范为准则，在地形图库、现状管线库、规划管线库、规划道路红线库等地图数据库的基础上，通过计算机及系统实现对设计信息的处理，完成管线设计计算、分析、绘图及方案的比较，从技术上避免规划管线与现状管线的矛盾和重复设计。

因此，应用型地理信息系统的功能不同于开发平台的基本功能，具有自己的特殊性。但是，应用系统的这些特殊功能主要应该依靠基础 GIS 提供的基本功能来开发和实现。

(7) 经费和管理。由于系统开发是一项复杂的系统工程，为保证系统开发工作的顺利进行，必须拟定好系统开发计划、系统管理措施、投资经费概算以及最后应提交的成果等。

2. 详细设计

详细设计又称为实际设计，其任务是根据总体设计方案确定的目标和阶段开发计划，紧密结合特定的硬件、基础软件和规范标准，进行子系统和数据库的详细设计，用于具体指导系统的开发。

详细设计的主要内容包括：

(1) 子系统设计。子系统设计以对用户需求的进一步详细调查为依据，分别完成各个子系统的逻辑结构设计、数据库设计、功能模块设计、用户界面设计等。每个子系统设计的内容大体类似于总体设计的内容，但应更加详细和具体，作为各个子系统实施的指导性文件。

(2) 数据库设计。主要内容包括：数据源的分析与选择；数据分类与分层的确定；数据获取方案的规定；数据编码设计；实体属性表与属性关系的设计；属性数据类型的建立；数据质量标准的规定；地理定位控制的确定及其他有关问题的

规定等。

(3) 功能模块设计。详细描述各功能模块的内容,实现的技术和算法,输入输出的数据项和格式等。

(4) 用户界面设计。用户界面是人机对话的工具,它与功能模块一一对应,做到各模块之间界面的形式一致,相同功能要用相同的图标显示。界面可以分为若干层,便于逐层调用。根据功能模块的不同,可以分别采用菜单式、命令式或表格式的界面。所有界面应体现以人为本的原则,做到界面友好、美观,并随时提供丰富的帮助信息,使用户易懂、易学、易掌握。

2.3　系统实施

系统实施是在系统设计的原则指导下,按照详细设计方案确定的目标、内容和方法,分阶段、分步骤完成系统开发的过程。

系统实施的内容包括:

(1) 系统硬件和软件的引进及调试。其实施的步骤如图 7－7 所示:

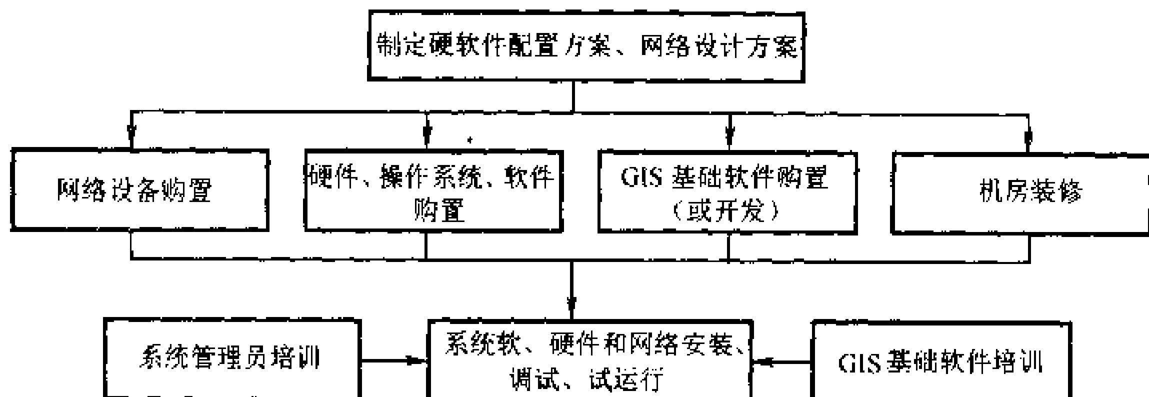


图 7－7　系统硬件、软件引进实施步骤

(据魏绥臣等,1998)

(2) 系统数据库建立。其实施内容包括数据源的选择、数据源的现势更新和处理、数据格式的定义和转换、数据采集方法的确定、数据编辑处理、数据质量控制、建立数据库实体等。

(3) 应用管理系统的开发。指在地理信息系统基础软件的基础上进行二次开发,建立应用管理系统。内容包括利用基础软件提供的开发语言进行编程、以各种菜单形式建立用户应用界面、应用模块的开发、建立图形符号库、编写用户操作手册等。

(4) 系统测试和联调。对系统开发完成的每一个模块,均应进行测试。将模块组装成系统时,也应进行联调和测试。系统测试是指利用人工或自动的方法测试和评价各个模块,验证模块是否满足规定的要求,检查设计指标与实际结